

Zero-point quantum fluctuations in cosmology

Lukas Hollenstein, Maud Jaccard, Michele Maggiore, and Ermis Mitsou
*Département de Physique Théorique and Center for Astroparticle Physics,
Université de Genève, 24 quai Ansermet, CH-1211 Genève 4, Switzerland**

We re-examine the classic problem of the renormalization of zero-point quantum fluctuations in a Friedmann-Robertson-Walker background. We discuss a number of issues that arise when regularizing the theory with a momentum-space cutoff, and show explicitly how introducing non-covariant counter-terms allows to obtain covariant results for the renormalized vacuum energy-momentum tensor. We clarify some confusion in the literature concerning the equation of state of vacuum fluctuations. Further, we point out that the general structure of the effective action becomes richer if the theory contains a scalar field ϕ with mass m smaller than the Hubble parameter $H(t)$. Such an ultra-light particle cannot be integrated out completely to get the effective action. Apart from the volume term and the Einstein-Hilbert term, that are reabsorbed into renormalizations of the cosmological constant and Newton's constant, the effective action in general also has a term proportional to $F(\phi)R$, for some function $F(\phi)$. As a result, vacuum fluctuations of ultra-light scalar fields naturally lead to models where the dark energy density has the form $\rho_{\text{DE}}(t) = \rho_X(t) + \rho_Z(t)$, where ρ_X is the component that accelerates the Hubble expansion at late times and $\rho_Z(t)$ is an extra contribution proportional to $H^2(t)$. We perform a detailed comparison of such models with CMB, SNIa and BAO data.

I. INTRODUCTION

The computation of the expectation value of the energy-momentum tensor in curved space, and particularly in a Friedmann-Robertson-Walker (FRW) background, is a classic and much studied problem that has important implications for understanding the origin of dark energy. The renormalization of $\langle 0|T_{\mu\nu}|0\rangle$ in a FRW background can be performed either with covariant regularization schemes, such as dimensional regularization or point-splitting [1], or using a cutoff Λ_c in momentum space [2–4]. The former option has the advantage of preserving general covariance explicitly, while a momentum-space cutoff can be more intuitive, for instance when developing naturalness arguments. The two schemes should give the same results for the renormalized quantities. However, the explicit breaking of general covariance by a momentum-space cutoff leads to some technical subtleties, that apparently have given rise to some confusion in both the old and the recent literature. For instance it was claimed that regularizing a theory with a sharp momentum cutoff Λ_c in a FRW background leads to pathologies [4] connected to the fact that the resulting vacuum energy density and pressure satisfy an equation of state (EOS) that is not consistent with general covariance [5]. It was also claimed that energy-momentum conservation is violated by the flow of momentum modes across the cutoff, induced by the cosmological red-shift [6–8]. There even appears to be some confusion as to what general covariance implies for $\langle 0|T_{\mu\nu}|0\rangle$. For a constant term such as the Λ_c^4 diver-

gence, general covariance implies that its contribution to $\langle 0|T_{\mu\nu}|0\rangle$ must be proportional to the metric $g_{\mu\nu}$, so its EOS parameter is $w = -1$. However, contrary to some statements in the literature (e.g. [8, 9]) this is not true for a time-dependent term such as the quadratic divergence $H^2(t)\Lambda_c^2$, where $H(t)$ is the Hubble parameter. We will show that general covariance implies that the contribution of this term to $\langle 0|T_{\mu\nu}|0\rangle$ must be proportional to the Einstein tensor $G_{\mu\nu}$. The corresponding EOS parameter is time-dependent and equal to the total EOS parameter of the background, $w_{\text{tot}}(t) = p_{\text{tot}}(t)/\rho_{\text{tot}}(t)$, rather than being $w = -1$.

The aim of this paper is two-fold. First, we will clarify the above technical issues. When using a non-covariant regularization scheme one must also include non-covariant counter-terms. We will see explicitly how introducing the appropriate non-covariant counter-terms allows us to obtain fully covariant results for the renormalized quantities. As a consequence, no pathologies arise even when employing a sharp momentum cutoff. The renormalized energy density and pressure satisfy the standard energy-momentum conservation. The EOS parameters associated to the terms arising from both the quartic and the quadratic divergences are dictated by general covariance and fully agree with the results obtained within covariant regularization schemes. We will also discuss a possible alternative to the standard renormalization of vacuum energy, in which the subtraction of the flat-space contribution from the vacuum energy computed in FRW is performed already at the purely classical level, invoking arguments from the ADM formalism of General Relativity, as well as from AdS/CFT correspondence, developing an argument that we presented in [10, 11].

After having shown that no new physics appear merely as a consequence of employing a non-covariant regularization scheme, we will turn to the second aim of this paper,

*Electronic address: lukas.hollenstein@unige.ch, maud.jaccard@unige.ch, michele.maggiore@unige.ch, ermis.mitsou@unige.ch

which is to show that, in contrast, the general structure of $\langle 0|T_{\mu\nu}|0\rangle$ changes in an interesting manner if the theory contains a scalar field ϕ whose mass m is smaller than the Hubble parameter $H(t)$ at the time of interest. In particular, this applies to an ultra-light scalar field with $m < H_0$. In this case the vacuum expectation value (VEV) of the energy-momentum tensor $T_{\mu\nu}$ acquires extra terms, compared to the standard case where all matter fields are very massive compared to $H(t)$. In general the VEV of $T_{\mu\nu}$ can be obtained by functional differentiation,

$$\langle 0|T_{\mu\nu}|0\rangle = -\frac{2}{\sqrt{-g}} \frac{\delta S_{\text{eff}}}{\delta g^{\mu\nu}}, \quad (1)$$

where S_{eff} is the effective action for gravity, which is obtained by treating the metric $g_{\mu\nu}$ as a classical background and integrating out the massive matter degrees of freedom (see e.g. refs. [12–15]). One can perform the calculation of the effective action using manifestly covariant regularization schemes such as dimensional regularization or point-splitting. The effective action is then explicitly generally covariant and depends only on structures such as the volume term $\int d^4x \sqrt{-g}$, the Einstein-Hilbert term $\int d^4x \sqrt{-g} R$, and higher-derivative terms such as $\int d^4x \sqrt{-g} R_{\mu\nu} R^{\mu\nu}$. The volume term is reabsorbed into a renormalization of the cosmological constant (giving rise to the cosmological constant naturalness problem) while the term in the effective action proportional to the Einstein-Hilbert term renormalizes Newton’s constant. Higher-derivative terms such as $R_{\mu\nu} R^{\mu\nu}$ can be seen as genuinely new effects due to vacuum fluctuations, but they can only be relevant when the curvature radius is of order of the Planck length. In particular, their impact is utterly negligible in the recent cosmological epoch.

The basic idea we develop in this paper is that if a scalar field ϕ with mass $m < H_0$ is present it cannot be integrated out completely, since it is not massive with respect to the energy scale of interest in cosmology, $H(t)$. Thus, after integrating over the massive degrees of freedom, the effective action will be a functional of both the metric $g_{\mu\nu}$ and the field ϕ (or, more precisely, of its low-frequency modes, that we will denote collectively as χ). This allows for a richer structure of the effective action and therefore of the corresponding $\langle 0|T_{\mu\nu}|0\rangle$. In particular, the effective action will generically include a term $\propto \int d^4x \sqrt{-g} f(\chi) R$, with some function $f(\chi)$. Even at low curvatures such a term is not suppressed with respect to the Einstein-Hilbert term, and can have interesting phenomenological consequences. We will see that this indeed leads to an explicit realization of a model of interacting vacuum fluctuations recently proposed in refs. [10, 11], where the dark energy tracks the total energy density of the Universe.

The paper is organized as follows. In section II we discuss the renormalization of vacuum energy in FRW using a momentum-space cutoff. In section III we discuss the effect of an ultra-light scalar field on $\langle 0|T_{\mu\nu}|0\rangle$. We then construct a cosmological model based on these considera-

tions and we analyze its phenomenological consequences.

We use signature $(-+++)$ and natural units where $\hbar = c = 1$, so $G = M_{\text{Pl}}^{-2}$, where M_{Pl} is the Planck mass. When we specialize our results to FRW, we work in a spatially flat FRW metric with cosmic time t , scale factor $a(t)$ and Hubble parameter $H(t) = (1/a)da/dt = \dot{a}/a$. Today, the Hubble parameter and the critical density take the values H_0 and $\rho_0 = 3H_0^2/(8\pi G)$, respectively.

II. VACUUM ENERGY AND RENORMALIZATION IN FRW SPACE-TIME

A. Renormalization and general covariance with a momentum-space cutoff

To illustrate the kind of problems that arise when computing $\langle 0|T_{\mu\nu}|0\rangle$ with a cutoff Λ_c over momenta, let us begin with the Λ_c^4 divergence in flat space. Considering for instance a single massless real scalar field, a straightforward calculation gives

$$\rho_{\text{bare}}(\Lambda_c) \equiv \langle 0|T_{00}|0\rangle = \frac{\Lambda_c^4}{16\pi^2}, \quad (2)$$

where the subscript “bare” emphasizes that this is still a bare, cutoff-dependent quantity, rather than the physical, renormalized energy density. For the pressure, which is given by $p = (1/3) \sum_i \langle 0|T_{ii}|0\rangle$, an analogous computation gives [4, 16] $p_{\text{bare}}(\Lambda_c) = \Lambda_c^4/(48\pi^2)$, and therefore

$$p_{\text{bare}}(\Lambda_c) = \frac{1}{3} \rho_{\text{bare}}(\Lambda_c). \quad (3)$$

This result has sometimes lead to the interpretation that the EOS of vacuum fluctuation is $p = w\rho$ with $w = 1/3$ (see e.g. footnote 19 of ref. [5]). In contrast, using a scheme such as dimensional regularization that preserves Lorentz invariance, in flat space one automatically gets $\langle 0|T_{\mu\nu}|0\rangle \propto \eta_{\mu\nu}$ and therefore $p = -\rho$.

The resolution of this apparent discrepancy is that the energy density and pressure given in eqs. (2) and (3) are bare quantities rather than renormalized ones and, in a regularization scheme that breaks Lorentz invariance, the EOS parameter $w_{\text{bare}} \equiv p_{\text{bare}}/\rho_{\text{bare}}$ is not the same as the physical EOS parameter for the renormalized quantities, $w_{\text{ren}} \equiv p_{\text{ren}}/\rho_{\text{ren}}$. Indeed, regularizing the theory with a cutoff over momenta breaks the Lorentz symmetry of Minkowski space, since the notion of maximum spatial momentum is not invariant under boosts. This means that in this scheme the renormalization procedure also involves counter-terms that are not Lorentz invariant. To give an explicit example of such a counter-term it is convenient to work with a generic metric $g_{\mu\nu}$ at first, before setting $g_{\mu\nu} = \eta_{\mu\nu}$ at the end. This allows us to derive the vacuum energy-momentum tensor from $\langle 0|T_{\mu\nu}|0\rangle = -(2/\sqrt{-g}) \delta S/\delta g^{\mu\nu}$, a procedure that facilitates the extension to curved space-time. The standard

Lorentz-invariant counter-term in the action is proportional to the volume term

$$S_A = A(\Lambda_c) \int d^4x \sqrt{-g}, \quad (4)$$

with $A(\Lambda_c)$ a suitably chosen function of the cutoff. An example of a Lorentz-breaking counter-term is

$$S_B = B(\Lambda_c) \int d^4x \sqrt{-g} g^{00}. \quad (5)$$

Using $\delta(\sqrt{-g}) = -(1/2)\sqrt{-g} g_{\mu\nu} \delta g^{\mu\nu}$ we see that the contribution to $\langle 0|T_{\mu\nu}|0\rangle$ from S_A is

$$-\frac{2}{\sqrt{-g}} \frac{\delta S_A}{\delta g^{\mu\nu}} = A(\Lambda_c) g_{\mu\nu}, \quad (6)$$

while the contribution from S_B is

$$-\frac{2}{\sqrt{-g}} \frac{\delta S_B}{\delta g^{\mu\nu}} = B(\Lambda_c) (g^{00} g_{\mu\nu} - 2\delta_\mu^0 \delta_\nu^0). \quad (7)$$

Evaluating them on the flat metric $g_{\mu\nu} = \eta_{\mu\nu}$ and using the flat-space expression $T_{\mu\nu} = (\rho, p, p, p)$ (or, more generally, using the FRW metric $g_{\mu\nu} = (-1, a^2, a^2, a^2)$ and $T_{\mu\nu} = (\rho, a^2 p, a^2 p, a^2 p)$), we see that S_A produces a counter-term for the energy density given by $\rho_A = -A(\Lambda_c)$ and a counter-term $p_A = -\rho_A = A(\Lambda_c)$ for the pressure. In contrast S_B produces the counter-terms $\rho_B = +p_B = -B(\Lambda_c)$.

For illustration we consider again a minimally coupled massless scalar field where $\rho_{\text{bare}}(\Lambda_c) = \Lambda_c^4/(16\pi^2)$ and $p_{\text{bare}}(\Lambda_c) = \Lambda_c^4/(48\pi^2)$. After adding these counter-terms to the bare Λ_c^4 divergences we get

$$\begin{aligned} \rho_{\text{ren}} &= \rho_{\text{bare}}(\Lambda_c) + \rho_{\text{count}}(\Lambda_c) \\ &= \frac{\Lambda_c^4}{16\pi^2} - A(\Lambda_c) - B(\Lambda_c), \end{aligned} \quad (8)$$

$$\begin{aligned} p_{\text{ren}} &= p_{\text{bare}}(\Lambda_c) + p_{\text{count}}(\Lambda_c) \\ &= \frac{\Lambda_c^4}{48\pi^2} + A(\Lambda_c) - B(\Lambda_c). \end{aligned} \quad (9)$$

We can now choose $A(\Lambda_c)$ and $B(\Lambda_c)$ such that $A + B = \Lambda_c^4/(16\pi^2) - \rho_{\text{finite}}$ (where ρ_{finite} is a finite part) and $A - B = -\Lambda_c^4/(48\pi^2) - \rho_{\text{finite}}$, i.e. $A(\Lambda_c) = \Lambda_c^4/(48\pi^2) - \rho_{\text{finite}}$ and $B(\Lambda_c) = 2\Lambda_c^4/(48\pi^2)$. In this way we get $\rho_{\text{ren}} = \rho_{\text{finite}}$ and $p_{\text{ren}} = -\rho_{\text{finite}}$, and therefore the renormalized quantities satisfy $p_{\text{ren}} = -\rho_{\text{ren}}$. Thus, Lorentz invariance can be restored for the renormalized quantities despite the fact that it was broken by the regularization scheme in the first place.

Having understood this simple point in the flat-space case, we can apply a similar strategy in curved space, except that now the guiding symmetry principle is no longer Lorentz invariance, but rather general covariance. In curved space one can again use regularization schemes, such as dimensional regularization or point-splitting, that preserve general covariance explicitly, and in this case one finds a fully covariant result already at the level of the

bare quantities. In particular, $\langle 0|T_{\mu\nu}|0\rangle$ can be written as in eq. (1), where S_{eff} is made of coordinate-invariant quantities. For instance, for a minimally coupled scalar field in $d = 4 - \epsilon$ dimensions, the UV divergent part of the effective action has the form (see e.g. eqs. (6.44)-(6.49) of ref. [1])

$$S_{\text{eff}} = \int d^4x \sqrt{-g} [c_1(\epsilon) + c_2(\epsilon) R + c_3(\epsilon) (R_{\mu\nu\rho\sigma} R^{\mu\nu\rho\sigma} - R_{\mu\nu} R^{\mu\nu} - 6\Box R)], \quad (10)$$

where $c_1(\epsilon), c_2(\epsilon), c_3(\epsilon)$ contain a term diverging as $1/\epsilon$ and a finite part. It is interesting to compare this with the structure of divergences obtained using a cutoff Λ_c over momenta, and specializing for simplicity directly to a FRW background. In this case, considering for definiteness a massless and minimally coupled real scalar field, we get $(\langle 0|T_{\mu\nu}|0\rangle)_{\text{bare}} = \text{diag}(-\rho_{\text{bare}}, p_{\text{bare}}, p_{\text{bare}}, p_{\text{bare}})$, where [4]

$$\rho_{\text{bare}}(\Lambda_c) = \frac{\Lambda_c^4}{16\pi^2} + \frac{H^2(t)\Lambda_c^2}{16\pi^2} + \mathcal{O}(H^4 \ln \Lambda_c), \quad (11)$$

$$p_{\text{bare}}(\Lambda_c) = \frac{\Lambda_c^4}{48\pi^2} + c_1 \frac{H^2(t)\Lambda_c^2}{16\pi^2} + \mathcal{O}(H^4 \ln \Lambda_c). \quad (12)$$

The coefficient c_1 depends on the cosmological epoch: $c_1 = -1/3$ during De Sitter (DS), $c_1 = +1$ during radiation dominance (RD) and $c_1 = 2/3$ during matter dominance (MD) [10].

The Λ_c^4 divergence is of course the same as in flat space. Again, it cannot be renormalized simply using the covariant counter-term (4), since the coefficient of the Λ_c^4 term in the pressure is not equal to minus the coefficient of the Λ_c^4 term in the energy density. Rather, one must also introduce the counter-term (5) and adjust these two counter-terms so that the corresponding renormalized energy density and pressure satisfy $p_{\text{ren}} = -\rho_{\text{ren}}$. An equivalent way to phrase this point is to observe that the Λ_c^4 terms in $\rho_{\text{bare}}(\Lambda_c)$ and $p_{\text{bare}}(\Lambda_c)$ do not satisfy the continuity equation, i.e.

$$\dot{\rho}_{\text{bare}} + 3H(\rho_{\text{bare}} + p_{\text{bare}}) \neq 0, \quad (13)$$

since for these quartically divergent terms $\dot{\rho}_{\text{bare}} = 0$ but $\rho_{\text{bare}} + p_{\text{bare}} \neq 0$. This is just a consequence of the fact that this regularization scheme breaks general covariance. Therefore energy-momentum conservation, which is a consequence of general covariance, does not hold for the bare quantities. There is nothing wrong or unphysical about this, as long as we choose the counter-terms such that the *renormalized* energy density and pressure satisfy energy-momentum conservation, which they do, since we saw that we can renormalize the quartic divergence so that, for this term, $p_{\text{ren}} = -\rho_{\text{ren}}$.

Similar but somewhat more interesting considerations can be made for the quadratically divergent term. First of all, let us see what the EOS of the term $\propto H^2(t)\Lambda_c^2$ should be, in order to satisfy energy-momentum conservation. Writing $\rho(t) = cH^2(t)$, where c is a constant,

$p(t) = w(t)cH^2(t)$, and plugging this into the conservation equation $\dot{\rho} + 3H(\rho + p) = 0$, we get

$$3[1 + w(t)] = -\frac{1}{H^3} \frac{dH^2}{dt} = -\frac{a}{H^2} \frac{dH^2}{da}, \quad (14)$$

Using the Friedmann equation $H^2(t) = (8\pi G/3) \sum_i \rho_i(t)$, where the index i runs through all contributions to the energy density of the Universe (matter, radiation, dark energy, etc.) and using $\rho_i(t) \propto a^{-3(1+w_i)}$, we get

$$w(t) = \frac{\sum_i w_i \rho_i(t)}{\sum_i \rho_i(t)} \equiv w_{\text{tot}}(t). \quad (15)$$

This is simply the total EOS parameter $w_{\text{tot}}(t) = p_{\text{tot}}/\rho_{\text{tot}}$ that tracks the EOS parameter of the dominant component, so it evolves from $w_{\text{tot}}(t) \simeq 1/3$ during RD to $w_{\text{tot}}(t) \simeq 0$ during MD, and finally $w_{\text{tot}}(t) \simeq -1$ if there is a dark energy dominated phase with $w_{\text{DE}} \simeq -1$.

We can now see that the terms $\propto H^2(t)\Lambda_c^2$ in eqs. (11) and (12) do not satisfy energy-momentum conservation, since the bare EOS parameter c_1 is not equal to $w_{\text{tot}}(t)$. Rather, the values $c_1 = \{+1, +2/3, -1/3\}$ computed assuming a pure RD, MD and DS epoch, respectively, are reproduced by

$$c_1(t) = w_{\text{tot}}(t) + 2/3. \quad (16)$$

The situation is completely analogous to that discussed for the Λ_c^4 divergence. If the quadratically divergent terms in the bare energy density and pressure were satisfying energy-momentum conservation we could renormalize them with a covariant counter-term, which in this case would simply be the Einstein-Hilbert term $\int d^4x \sqrt{-g} R$, so they would just be reabsorbed into a renormalization of Newton's constant. The fact that the quadratically divergent terms in the bare energy density and pressure do not satisfy energy-momentum conservation means that we must also introduce a non-covariant counter-term, chosen so that the contribution proportional to $H^2(t)$ in the renormalized energy density and pressure satisfy $\rho_{\text{ren}}(t) = \text{const.} \times H^2(t)$ and $p_{\text{ren}}(t) = w_{\text{tot}}(t)\rho_{\text{ren}}(t)$. We explicitly construct the required non-covariant counter-term in appendix A.

To conclude this section, let us observe that a dark energy term $\propto H^2(t)M_{\text{Pl}}^2$ can be obtained rather generally, for instance from holographic considerations. In this context it is sometimes stated that, if the regularization of the divergent expressions is handled in a Lorentz-invariant manner, as with dimensional regularization, the corresponding VEV of the energy-momentum tensor will have the form $\langle 0|T_{\mu\nu}|0\rangle \propto g_{\mu\nu}$, i.e. $w = -1$ [9]. The discussion presented above shows that this is not correct. First of all, Lorentz invariance is irrelevant here as it is not a symmetry of the FRW space-time. Rather, the relevant symmetry is general covariance. As we have seen above, for the term $\propto H^2(t)$ in the energy density, general covariance fixes the EOS parameter to $w = w_{\text{tot}}(t)$

rather than to $w = -1$. The two results coincide only in a De Sitter epoch.

An equivalent way of understanding this point is to observe that a term $\propto H^2(t)$ in $\langle 0|T_{\mu\nu}|0\rangle$ is obtained by taking the functional derivative of the Einstein-Hilbert term in the effective action (and not of the volume term). Therefore its covariant form is proportional to the Einstein tensor $G_{\mu\nu}$ rather than to the metric $g_{\mu\nu}$ (see also the explicit covariant computation in [17]). Again, if $H(t)$ is constant in time $G_{\mu\nu} \propto g_{\mu\nu}$ and the two results agree, otherwise they are different. This is also the most direct way of deriving the EOS for the term $\propto H^2(t)$ in $\langle 0|T_{\mu\nu}|0\rangle$: using the fact that this contribution to $\langle 0|T_{\mu\nu}|0\rangle$ is proportional to $G_{\mu\nu}$, together with Einstein equations $G_{\mu\nu} = (8\pi G)T_{\mu\nu}^{\text{tot}}$ we see that in the end this contribution to $\langle 0|T_{\mu\nu}|0\rangle$ is proportional to $T_{\mu\nu}^{\text{tot}}$. So, in FRW, the term proportional to $H^2(t)$ in the vacuum energy density and pressure satisfy $p_{\text{vac}}/\rho_{\text{vac}} = p_{\text{tot}}/\rho_{\text{tot}} \equiv w_{\text{tot}}(t)$.

B. Comparison with previous work

The fact that the terms $\propto H^2(t)\Lambda_c^2$ in $\rho_{\text{bare}}(\Lambda_c)$ and $p_{\text{bare}}(\Lambda_c)$ do not respect energy-momentum conservation appears to have generated some recurrent confusion in the literature. Renormalization of the energy-momentum tensor with a momentum-space cutoff in FRW was discussed long ago in a classic paper by Fulling and Parker [4]. However, we believe that our discussion above sheds light on some aspects of ref. [4]. In that paper, the authors considered the $\langle 00 \rangle$ and the $\langle 0i \rangle$ components of the Einstein equations in a FRW space-time, both written in terms of the bare quantities¹

$$\left(\frac{\dot{a}}{a}\right)^2 - \frac{1}{3}\Lambda_0 = \frac{8\pi G_0}{3}\rho_{\text{bare}}, \quad (17)$$

$$2\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 - \Lambda_0 = -8\pi G_0 p_{\text{bare}}, \quad (18)$$

where Λ_0 is the bare cosmological constant, not to be confused with the cutoff Λ_c , and G_0 the bare Newton's constant. Regularizing the theory with a sharp cutoff in momentum space, and restricting for simplicity to the contribution of a minimally-coupled massless scalar field in De Sitter space, they find that both ρ_{bare} and p_{bare} have a quartic as well as a quadratic divergence, given by eqs. (11) and (12), with $c_1 = -1/3$ for De Sitter. Including also the finite parts ρ_{finite} and p_{finite} we therefore

¹ Compared to eq. (5.4) of ref. [4], for notational simplicity we limit to the spatially flat case and we do not explicitly write the term which is related to the renormalization of terms quadratic in the Riemann tensor, denoted $H_{\mu\nu}$ in ref. [4].

have

$$\rho_{\text{bare}}(\Lambda_c) = \frac{\Lambda_c^4}{16\pi^2} + \frac{H^2\Lambda_c^2}{16\pi^2} + \mathcal{O}(\ln \Lambda_c) + \rho_{\text{finite}}, \quad (19)$$

$$p_{\text{bare}}(\Lambda_c) = \frac{\Lambda_c^4}{48\pi^2} - \frac{H^2\Lambda_c^2}{48\pi^2} + \mathcal{O}(\ln \Lambda_c) + p_{\text{finite}}. \quad (20)$$

Thus eq. (17) can be rewritten as

$$\left(\frac{\dot{a}}{a}\right)^2 - \frac{1}{3} \left(\frac{\Lambda_0 + \frac{G_0\Lambda_c^4}{2\pi}}{1 - \frac{G_0\Lambda_c^2}{6\pi}} \right) = \frac{8\pi}{3} \left(\frac{G_0}{1 - \frac{G_0\Lambda_c^2}{6\pi}} \right) \rho_{\text{finite}} \quad (21)$$

which shows that, at least in this equation, the quartic divergence is reabsorbed into a renormalization of the cosmological constant and the quadratic divergence into a renormalization of Newton's constant.²

The situation is however different in eq. (18) since, if we now insert the explicit expression of p_{bare} , we see that in this equation neither the quartic nor the quadratic divergence can be reabsorbed into the above renormalization of Λ_0 and G_0 . The authors of ref. [4] propose the following solution. They observe that eq. (18) can be obtained by taking a time derivative of eq. (17), upon use of the conservation equation $\dot{\rho}_{\text{bare}} + 3H(\rho_{\text{bare}} + p_{\text{bare}}) = 0$. However, as we saw above, ρ_{bare} and p_{bare} given in eqs. (19) and (20), do not satisfy this conservation equation, neither for the quartically divergent terms nor for the quadratically divergent terms. They then discard the results (19) and (20) arguing that a sharp cutoff in momentum space is not a correct regularization, and that one should rather evaluate ρ_{bare} using a smooth cutoff function, i.e. replacing

$$\int_0^{a\Lambda_c} dk \rightarrow \int_0^\infty dk f(k/a\Lambda_c), \quad (22)$$

where $f(k/a\Lambda_c)$ is a smoothed version of the theta function $\theta(1 - k/a\Lambda_c)$. (Here k denotes the comoving momentum, while in terms of the physical momentum, $k_{\text{phys}} = k/a(t)$, the cutoff is at Λ_c .) Further, they argue that p_{bare} should be *defined* by $\dot{\rho}_{\text{bare}} + 3H(\rho_{\text{bare}} + p_{\text{bare}}) \equiv 0$.³ In summary, according to ref. [4] the bare energy density should be computed as an integral over modes with a cutoff given by a smoothed theta function, rather than with a sharp cutoff, while the bare pressure, instead of being given by the corresponding integral over modes with a

smoothed theta function, should be *defined* by the conservation equation.

Our discussion in the previous subsection however allows us to get a better understanding of this issue. First of all, there is nothing wrong with using a sharp momentum-space cutoff, and physical results should not depend on the choice of the regulator. The fact that it is impossible to reabsorb the divergences in eqs. (17) and (18) into a renormalization of Newton's constant (by means of a counter-term proportional to the Einstein-Hilbert term) is not a sign of a pathology. Instead, it is a direct consequence of the need for non-covariant counter-terms to restore the symmetry after regularizing with a scheme that breaks general covariance. The Fulling-Parker prescription, which consists of using eq. (22) and defining $p_{\text{bare}}(\Lambda_c)$ from $\dot{\rho}_{\text{bare}} + 3H(\rho_{\text{bare}} + p_{\text{bare}}) = 0$, can be seen as the definition of a regularization scheme that tries to enforce energy-momentum conservation by hand at the level of the bare quantities even in the presence of a momentum-space regularization. In this sense it is a legitimate, but certainly unnecessary (and somewhat awkward) choice of regularization.⁴

A certain confusion seems to exist even in the more recent literature. When using a momentum-space cutoff it is not always appreciated that, even if the bare quantities do not satisfy energy-momentum conservation,

$$\dot{\rho}_{\text{bare}} + 3H(\rho_{\text{bare}} + p_{\text{bare}}) \neq 0, \quad (23)$$

it is still possible to renormalize the theory such that the renormalized quantities satisfy it,

$$\dot{\rho}_{\text{ren}} + 3H(\rho_{\text{ren}} + p_{\text{ren}}) = 0. \quad (24)$$

Equivalently, when the regularization scheme breaks general covariance one needs to introduce non-covariant counter-terms and as a consequence the bare EOS parameter $w_{\text{bare}} \equiv p_{\text{bare}}/\rho_{\text{bare}}$ is different from the physical EOS parameter $w \equiv p_{\text{ren}}/\rho_{\text{ren}}$. For instance, in ref. [7] the fact that the bare vacuum energy-momentum tensor does not satisfy the conservation equation leads the author to propose a modification of the Einstein equations of the form $G_{\mu\nu} + \Sigma_{\mu\nu} = 8\pi GT_{\mu\nu}$, where $\Sigma_{\mu\nu}$ is a tensor that satisfies $\nabla_\mu \Sigma^{\mu\nu} = 8\pi G \nabla_\mu T^{\mu\nu} \neq 0$. However, once again, there is no special physical meaning in the fact that the *bare* energy-momentum tensor

² The logarithmically divergent terms, that we have not written explicitly, are reabsorbed into the renormalization of higher derivative operators such as $R_{\mu\nu}R^{\mu\nu}$ [4].

³ Actually, they write $p_{\text{bare}} = \mathcal{D}\rho_{\text{bare}} \equiv (3a^2\dot{a})^{-1}d(a^3\rho_{\text{bare}})/dt$, see eqs. (5.5) and (5.7) of ref. [4], which is the same as $\dot{\rho}_{\text{bare}} + 3H(\rho_{\text{bare}} + p_{\text{bare}}) = 0$. The role of the smoothing function $f(k/a\Lambda_c)$ is that the time derivative in the operator \mathcal{D} acts also on it, because of the factor $a(t)$ in the argument of f , generating a term proportional to f' , without which it would not be possible to impose eq. (13) to be satisfied.

⁴ It should also be observed that general covariance is not fully recovered at the level of bare quantities even with this prescription. This clarifies another point of ref. [4] that the authors themselves considered disturbing, namely the fact that some convergent terms in the expression for the bare energy density of order m^2T^{-4} , where m is the mass of the scalar field and T the parameter of adiabatic regularization, do not form the (00) component of any divergenceless tensor, see the discussion at the bottom of page 186 of [4]. Once again, the point is that if the regularization scheme breaks general covariance, the bare quantities will not respect it either. On the other hand, a generally covariant result for the renormalized quantities can be obtained upon use of non-invariant counter-terms.

is not conserved. What appears on the right-hand side of the Einstein equations really is the conserved, physical energy-momentum tensor constructed from the renormalized vacuum energy density and pressure.

Similarly, in ref. [8] the EOS of the term $H^2(t)\Lambda_c^2$ was investigated and it was claimed that the EOS parameter for vacuum fluctuations is $w = -1 + \Lambda_c^2/(9\pi M_{\text{Pl}}^2)$. However, this is a statement about the ratio $p_{\text{bare}}(\Lambda_c)/\rho_{\text{bare}}(\Lambda_c)$ and not about the ratio of the corresponding renormalized quantities.⁵ Similar comments hold for the discussion of the EOS in [18].

The same criticism applies if one takes an effective field theory point of view rather than working in the context of full renormalization theory. Let us consider the theory which has been used to derive the result $\rho_{\text{bare}}(\Lambda_c) = \Lambda_c^4/(16\pi^2)$ as the low-energy limit of a more fundamental theory that takes over at a scale M . The total physical vacuum energy density will be given by the integral over the modes with $k < M$, that simply gives $M^4/(16\pi^2)$, plus the contribution from the modes with $k > M$ which can only be computed using the UV completion of the theory and will be of order M^4 , but comes with a numerical coefficient that cannot be computed with only the low-energy theory at hand. Thus, the exact coefficients of Λ_c^4 in the energy density and the pressure are unpredictable within an effective field theory and, therefore, the physical EOS parameter cannot be computed either. Thus, either one fixes the EOS parameter by means of symmetry principles, i.e. by demanding Lorentz invariance for the Λ_c^4 term or general covariance for the $H^2\Lambda_c^2$ term, or one simply has no handle on it. In any case, the EOS parameter derived from the ratio $p_{\text{bare}}(\Lambda_c)/\rho_{\text{bare}}(\Lambda_c)$ has no physical meaning.

C. An alternative to cosmological constant renormalization: subtractions in classical GR

The renormalization of the Λ_c^4 divergence gives rise to the infamous cosmological constant naturalness problem: the divergence in the bare vacuum energy, $\rho_{\text{bare}}(\Lambda_c) \equiv \langle 0|T_{00}|0\rangle$, is canceled by a counter-term $\rho_{\text{count}}(\Lambda_c)$ that corresponds to a counter-term proportional to $\int d^4x \sqrt{-g}$ in the action. The physical, renormalized vacuum energy is then $\rho_{\text{ren}} = \rho_{\text{bare}}(\Lambda_c) + \rho_{\text{count}}(\Lambda_c)$. The naturalness problem arises from the fact that $\rho_{\text{count}}(\Lambda_c)$ must cancel the Λ_c^4 divergence in $\rho_{\text{bare}}(\Lambda_c)$, leaving a finite part that is exceedingly small compared to Λ_c^4 , for all values of the cutoff larger than, say, the TeV scale, where quantum field theory is well tested.

At first sight the necessity of an extremely fine-tuned cancellation between $\rho_{\text{bare}}(\Lambda_c)$ and $\rho_{\text{count}}(\Lambda_c)$ can be dis-

turbing. It should however be observed that neither the bare term nor the counter-term have any physical meaning and only their sum is physical, so this fine tuning is different from an implausible cancellation between observable quantities. The same kind of cancellation appears for instance in the Casimir effect, and this might provide a first hint of what could be the correct treatment of vacuum energies in cosmology. In the Casimir effect one computes the vacuum energy density of a field in a finite volume (e.g. the electromagnetic field between two parallel conducting plates at a distance L) and subtracts from it the vacuum energy density computed in an infinite volume. If we regularize with a cutoff Λ_c over momenta both terms diverge as Λ_c^4 , but their difference is finite and depends only on the macroscopic scale L . Thus the quantity of interest is the difference between the vacuum energy in the given geometry and the vacuum energy in a reference geometry, which is just flat space-time in an infinite volume. This might suggest that, to obtain the physical effect of the vacuum energy density on the expansion of the Universe, one should analogously compute the vacuum energy density in a FRW space-time and subtract from it the value computed in a reference space-time, which could be naturally taken as Minkowski space. This procedure leads to a sort of “cosmological Casimir effect”.

An immediate objection to this analogy could be that in special relativity the zero of the energy can be chosen arbitrarily and only energy differences with respect to the ground state are relevant. In contrast, in GR we cannot choose the zero of the energy arbitrarily. One typically expects that “every form of energy gravitates”, such that the contribution of Minkowski space cannot simply be dropped. While it is certainly true that in GR the choice of the zero of the energy is not arbitrary, the point that we wish to make here is that the correct choice can be a non-trivial issue (see also [10, 11]). As a first example, consider the definition of energy in GR for asymptotically flat metrics. To carefully define the energy associated with a given field configuration it is convenient to use the Hamiltonian formulation of GR, which goes back to the classic paper by Arnowitt, Deser and Misner (ADM) [19] (see e.g. the textbook [20] for a very clear recent review). In order to define the Hamiltonian of GR one must first work in a finite three-dimensional volume, where it takes the form

$$H_{\text{GR}} = H_{\text{bulk}} + H_{\text{boundary}}. \quad (25)$$

The bulk term H_{bulk} is given by an integral over the three-dimensional finite spatial volume at fixed time, and H_{boundary} by an integral over its two-dimensional boundary. If we try to define the energy of a classical field configuration as the value of this Hamiltonian evaluated on the classical solution we encounter a problem: as a consequence of the invariance under diffeomorphisms, the volume term H_{bulk} vanishes when evaluated on any classical solution of the equations of motion. Then the whole energy comes from the boundary term. However,

⁵ Furthermore, the derivation in ref. [8] assumed that the EOS parameter consistent with general covariance is $w = -1$ while, as we saw above, for the term $\propto H^2(t)\Lambda_c^2$ general covariance gives $w = w_{\text{tot}}(t)$.

when the boundary term is evaluated on any asymptotically flat metric (including Minkowski) it diverges as the boundary is taken to infinity. The solution proposed by ADM is to subtract the same boundary term evaluated on Minkowski space-time. The resulting energy (or mass) is finite and provides the standard definition of mass in GR, known as the ADM mass. For instance, when applied to the Schwarzschild space-time, the ADM mass computed in this way turns out to be equal to the mass M that appears in the Schwarzschild metric. The ADM prescription can be summarized by saying that, in GR, the energy E associated to a classical asymptotically-flat space-time with metric $g_{\mu\nu}$ can be obtained from the Hamiltonian H_{GR} through

$$E = H_{\text{GR}}[g_{\mu\nu}] - H_{\text{GR}}[\eta_{\mu\nu}], \quad (26)$$

where $\eta_{\mu\nu}$ is the flat metric. Even if the context in which this formula is valid, namely asymptotically flat space-times, is different from the cosmological context that we are interested in here, eq. (26) still suffices to make the point that the intuitive understanding that GR requires any form of energy to act as a source for the gravitational field is not really correct. Equation (26) tells us that the energy associated to Minkowski space does not gravitate.

The ADM prescription has also been generalized to space-times that are not asymptotically flat, subtracting the contribution of some reference space-time whose boundary has the same induced metric as the geometry under consideration [21–24]. The issue has been studied in detail in particular for asymptotically AdS space-times, where the problem appears because the boundary term in the Hamiltonian diverges as the boundary used in its definition approaches the boundary of AdS. In this case it has been shown that the appropriate subtraction can be performed without even introducing a reference background, but by adding some local counter-terms to the boundary action. These are given by coordinate-invariant functionals of the intrinsic boundary geometry, fixed by the requirement to obtain a finite energy as the boundary is taken to infinity [25, 26]. This prescription is particularly appealing since for AdS the structure of the divergences of the boundary action is such that they can be removed by adding a finite polynomial in the boundary curvature and its derivatives. This is analogous to the fact that in QFT the UV divergences are removed by adding counter-terms that are polynomials of finite order in the fields. Indeed, in the context of the AdS/CFT correspondence, this way of removing divergences in the gravitational action on the AdS side corresponds to the renormalization of the UV divergences in the conformal quantum field theory that lives on the boundary [25–29]. This subtraction procedure correctly reproduces the masses of various known space-times that are asymptotically AdS. For instance, in $2+1$ dimensions it correctly gives the mass of the BTZ black hole (which is asymptotically AdS₃) and also reproduces the transformation law and conformal anomaly of the stress tensor in the dual CFT. In $3+1$ dimensions it correctly reproduces

the known value of the mass of the AdS₄-Schwarzschild solution and similar results can be obtained for asymptotically AdS₅ solutions. Furthermore, while this subtraction procedure provides a zero mass for pure AdS₄ space-times, it gives a non-vanishing value for the mass of pure AdS₅, which exactly matches the Casimir energy of the dual $\mathcal{N} = 4$ super Yang-Mills theory that lives on the global AdS₅ boundary with topology $S^3 \times R$ [25].

What we learn from these examples is that the intuitive argument that “GR requires that any form of energy contributes to the gravitational field” is in fact loose and generally incorrect. Rather, in all cases an appropriate subtraction, formulated either as the subtraction of the contribution of a reference space-time, or in terms of local counter-terms, is required already at the classical level in order to obtain an energy-momentum tensor which is well-defined and reproduces the known properties of the space-time under consideration.

From this vantage point it is rather natural to assume that the definition of energy associated to zero-point quantum fluctuations in a curved background should not escape this general rule. As we have seen, regularizing the theory with a cutoff Λ_c over comoving momenta, in FRW the bare energy density of zero-point fluctuations takes the form

$$[\rho_{\text{bare}}(\Lambda_c)]_{\text{FRW}} = [\rho_{\text{bare}}(\Lambda_c)]_{\text{Mink}} + \mathcal{O}(H^2(t)\Lambda_c^2) + \mathcal{O}(H^4(t)\ln\Lambda_c). \quad (27)$$

Here $[\rho_{\text{bare}}(\Lambda_c)]_{\text{Mink}}$ is the bare vacuum energy density in Minkowski space and, for a field of mass m , its general UV structure is

$$[\rho_{\text{bare}}(\Lambda_c)]_{\text{Mink}} = \mathcal{O}(\Lambda_c^4) + \mathcal{O}(m^2\Lambda_c^2) + \mathcal{O}(m^4\ln\Lambda_c) + \text{finite part}. \quad (28)$$

For example, for a real minimally coupled massless scalar field, $[\rho_{\text{bare}}(\Lambda_c)]_{\text{Mink}} = c_1\Lambda_c^4$ and

$$[\rho_{\text{bare}}(\Lambda_c)]_{\text{FRW}} = c_1\Lambda_c^4 + c_2H^2(t)\Lambda_c^2 + \mathcal{O}(H^4(t)\ln\Lambda_c), \quad (29)$$

with $c_1 = c_2 = 1/(16\pi^2)$.⁶ The quartic divergence is the same as that found in flat Minkowski space. This can be seen from the explicit computation and is of course more generally a consequence of the fact that the result must reduce to that obtained in flat space in the limit $H \rightarrow 0$.

Our discussion of the definition of energy in GR suggests the following alternative to the standard renormalization of the cosmological constant. Rather than

⁶ Actually, the value of c_2 also depends on the choice of modes, and hence of vacuum state, for the field under consideration. The value $c_2 = 1/(16\pi^2)$ is obtained using the modes that correspond to the usual Bunch-Davies vacuum. The most general result is obtained by taking a combination of these modes with arbitrary Bogoliubov coefficients. This changes the numerical value of the constant c_2 , but not the general structure of divergences displayed in eq. (27), see e.g. appendix B of [10].

eliminating the Λ_c^4 divergence with standard renormalization, which allows us to get rid of it at the price of the naturalness problem, another option is that we should first perform the appropriate subtraction, corresponding to the correct choice of the zero of the energy density. To understand what is the correct subtraction in FRW, we consider the Friedmann equation when on the right hand side of the Einstein equations we use as source $T_{\mu\nu} + \langle 0|T_{\mu\nu}|0\rangle$, where $T_\nu^\mu = \text{diag}(-\rho, p, p, p)$ is the ordinary classical contribution of matter, radiation, etc, and $\langle 0|T_{\mu\nu}|0\rangle$ is the corresponding contribution of zero-point fluctuations. The Friedmann equation then reads

$$H^2(t) = \frac{8\pi G}{3} (\rho + \langle 0|T_{00}|0\rangle). \quad (30)$$

We fix the subtraction procedure by requiring that Minkowski space (i.e. $H(t) = 0$) should be a solution in the limit $\rho \rightarrow 0$. This implies that all terms in $[\rho_{\text{vac}}]_{\text{FRW}}$ that do not vanish when $H \rightarrow 0$ must be subtracted, i.e. we must subtract the vacuum energy computed in Minkowski space. This eliminates the Λ_c^4 divergence (as well as other flat-space divergences such as terms $\propto m^2 \Lambda_c^2$ that appear for a massive field with mass m) in a way that does not suffer from the naturalness problem. The remaining divergences are renormalized in the standard way. In particular, the term $\propto H^2(t) \Lambda_c^2$ in the energy density is reabsorbed into the renormalization of Newton's constant (plus non-covariant counter-terms for renormalizing the pressure, as discussed in section II A). The renormalization of these remaining divergences does not suffer from any naturalness problem (and, in any case, they could not be affected by a constant shift in the zero-point of the energy).⁷

It is also interesting to observe that this procedure provides a new perspective for interpreting the results of ref. [30]. There the authors show that in a theory with a large number \mathcal{N} of quantum fields, because of the finite volume fluctuations of vacuum energy around its average value, Minkowski space would be unstable to black hole formation in regions of size $\mathcal{O}(\mathcal{N}^{1/4} l_{\text{Pl}})$, where l_{Pl} is the Planck length. This is of course a paradoxical result, and to escape this conclusion the authors of ref. [30] suggest that the quantum gravity length-scale for \mathcal{N} fields should be of order $\mathcal{N}^{1/4} l_{\text{Pl}}$ instead of simply l_{Pl} . In other words the UV cutoff should be parametrically smaller than the Planck mass, and of order $\mathcal{N}^{-1/4} M_{\text{Pl}}$. If this was the case, semiclassical gravity would break down at this scale, and the calculation used to prove black hole formation

would no longer be valid. With our proposal, in contrast, the problem is eliminated simply because vacuum fluctuations in Minkowski space do not gravitate.

By means of this approach it is also possible to understand why phase transitions in the early Universe do not produce an exceedingly large vacuum energy density [31]. Consider for instance a scalar field φ undergoing spontaneous symmetry breaking (SSB), with the typical double-well potential

$$V(\varphi) = -\frac{1}{2}\mu^2\varphi^2 + \frac{\lambda}{4}\varphi^4, \quad (31)$$

with $\mu^2 > 0$. Computing the energy density in the FRW background, one finds the vacuum energy associated to SSB to be

$$\rho_{\text{SSB}} = -\frac{\mu^4}{4\lambda} - \frac{\mu^2}{2\lambda}(\dot{H} + 2H^2). \quad (32)$$

The first term is the flat-space result that would typically give an unacceptably high value for the vacuum energy density. For instance, the electroweak transition would give $\rho_{\text{SSB}}^{1/4} = \mathcal{O}(100 \text{ GeV})$. With our proposal, however, this Minkowski space term does not contribute to the gravitating energy density, and one remains with the second term in eq. (32), that is $\mathcal{O}(\mu^2 H^2/\lambda)$ and therefore totally consistent with observations.

III. VACUUM ENERGY AND ULTRA-LIGHT SCALARS

The quadratic divergence proportional to $H^2(t) \Lambda_c^2$ has the interesting property that, if we set the cutoff Λ_c at a value of the order of the Planck mass M_{Pl} , it becomes of the order of the critical density of the Universe at time t , since $\rho_c(t) = (3/8\pi)H^2(t)M_{\text{Pl}}^2$. This is in sharp contrast with the quartic divergence Λ_c^4 , that for $\Lambda_c \sim M_{\text{Pl}}$ is more than 120 order of magnitudes larger than the observationally allowed value. Thus, the quadratic divergence might offer a hint for understanding the coincidence problem, namely why the vacuum energy density that drives the acceleration of the Universe is of the order of the critical density today. In this naive form, however, this argument does not lead us very far, since the quadratic divergence is simply reabsorbed into a renormalization of Newton's constant and does not leave observable effects. Still, the fact that a term proportional to $H^2(t)$ in the vacuum energy density might naturally be of the order of the critical density is quite intriguing and stimulates further investigation. In this section we will see that it is indeed possible to prevent the term $\propto H^2(t)$ of being reabsorbed into a renormalization of Newton's constant. However, to do so we are required to introduce an ultra-light degree of freedom into the theory. In this paper, for definiteness, we will consider a scalar field. However, most of our considerations in this section go through as well if we introduce an ultra-light massive graviton instead. The role of the scalar field is then played by the

⁷ Needless to say, none of the experiments that are usually mentioned as probes of vacuum fluctuations, such as the Casimir effect or the Lamb shift, are affected by such a choice of the zero point of the energy. These experiments, of course, do not probe directly the choice of the zero in the definition of the energy; rather they always probe energy differences, while the zero of the energy is defined, arbitrarily, by normal ordering the Hamiltonian.

helicity-0 component of the massive graviton. Ultra-light scalars with $m \sim H_0$ where first discussed, in the context of pseudo-Nambu-Goldstone bosons, in [32]. In ref. [33] it was first suggested that quantum effects from ultra-light scalar fields could lead to late-time acceleration.

A. Ultra-light scalars and the effective action

We begin by investigating the most general structure of the renormalized VEV of the energy-momentum tensor. First of all, to avoid misunderstandings, let us emphasize that the answer to this question does not depend on the regularization scheme employed. This may affect the structure of the *bare* $\langle 0|T_{\mu\nu}|0\rangle$, but not that of the renormalized one. In a theory without ultra-light scalar fields, a generally covariant regularization gives the most general structure of the bare $\langle 0|T_{\mu\nu}|0\rangle$ as in eqs. (1) and (10). In this case even the counter-terms are generally covariant, so the renormalized $\langle 0|T_{\mu\nu}|0\rangle$ can in general have a term proportional to $g_{\mu\nu}$ from the variation of the volume term (which is set to zero if one adopts the proposal in section II C, or otherwise reabsorbed into a renormalization of the cosmological constant). Further it can have a term proportional to the Einstein tensor $G_{\mu\nu}$ from the variation of the Einstein-Hilbert term, reabsorbed into Newton's constant, and contributions from the variation of the terms quadratic in the Riemann tensor. On dimensional grounds the latter are suppressed by a factor $(E/M_{\text{Pl}})^2$ with respect to the Einstein-Hilbert term at the energy (or curvature) scale E , and can therefore only be relevant at Planckian energies.

As we have discussed, using a momentum-space cut-off slightly complicates the analysis as the explicit calculation of the bare $\langle 0|T_{\mu\nu}|0\rangle$ in such a scheme produces non-covariant contributions. These are compensated by introducing non-covariant counter-terms, chosen so to restore general covariance for the renormalized quantities. Thus, in the end the renormalized $\langle 0|T_{\mu\nu}|0\rangle$ is independent of the regularization scheme, as it should. In the following, we will use a generally covariant regularization, so the general structure of the bare $\langle 0|T_{\mu\nu}|0\rangle$ is the same as that of the renormalized $\langle 0|T_{\mu\nu}|0\rangle$.

To understand if it is possible to obtain a more general result for $\langle 0|T_{\mu\nu}|0\rangle$ it is useful to recall that the VEV of $T_{\mu\nu}$ can be obtained by taking the functional derivative of the effective action with respect to the metric, see e.g. [14, 15]. The effective action $S_{\text{eff}}[g_{\mu\nu}]$ is derived from the matter action $S_m[g_{\mu\nu}, \psi]$ by integrating over the matter fields, here collectively denoted by ψ

$$e^{iS_{\text{eff}}[g_{\mu\nu}]} = \int \mathcal{D}\psi e^{iS_m[g_{\mu\nu}, \psi]}. \quad (33)$$

The matter energy-momentum tensor is defined as usual by $T_{\mu\nu} = -(2/\sqrt{-g}) \delta S_m / \delta g^{\mu\nu}$. In the path-integral formulation its vacuum expectation value is given by

$$\langle 0|T_{\mu\nu}|0\rangle = \frac{\int \mathcal{D}\psi T_{\mu\nu} e^{iS_m[g_{\mu\nu}, \psi]}}{\int \mathcal{D}\psi e^{iS_m[g_{\mu\nu}, \psi]}}, \quad (34)$$

and can be rewritten as

$$\begin{aligned} \langle 0|T_{\mu\nu}|0\rangle &= \frac{1}{e^{iS_{\text{eff}}[g_{\mu\nu}]}} \left(\frac{-2}{i\sqrt{-g}} \right) \frac{\delta}{\delta g^{\mu\nu}} e^{iS_{\text{eff}}[g_{\mu\nu}]} \\ &= -\frac{2}{\sqrt{-g}} \frac{\delta S_{\text{eff}}}{\delta g^{\mu\nu}}. \end{aligned} \quad (35)$$

The effective action defined in eq. (33) is the relevant quantity for describing the physics at an energy scale that is sufficiently low compared to the mass of all matter fields, so that the matter fields can be integrated out. In FRW the relevant scale is given by the Hubble parameter $H(t)$ at the time t of interest. Thus, one is implicitly assuming that all matter fields are very massive with respect to the Hubble scale, $m \gg H(t)$. However, if the theory includes a scalar field ϕ with mass $m \lesssim H(t)$, this assumption is violated since such a field cannot be integrated out completely. In particular, in the presence of an ultra-light scalar with $m \sim H_0$, the assumption is violated for all times $t \leq t_0$ (where t_0 denotes the present age of the Universe). More precisely, at time t we can integrate out the Fourier modes $\phi_{\mathbf{k}}$ (where \mathbf{k} denotes the comoving momentum, related to the physical momentum by $\mathbf{k}_{\text{phys}} = \mathbf{k}/a(t)$) whose frequency satisfies

$$\omega_k \equiv \sqrt{m^2 + (k^2/a^2)} \gg H(t). \quad (36)$$

The low-frequency modes with \mathbf{k} such that $\omega_k \lesssim H(t)$ remain in the effective action and should be treated as a classical background, on the same footing as the FRW metric $g_{\mu\nu}$.⁸ It is then convenient to split the field $\phi(x)$ into its low- and high-frequency parts. We first expand $\phi(x)$ in Fourier modes,

$$\phi(x) = \int \frac{d^3k}{(2\pi)^3 \sqrt{2k}} \left[a_{\mathbf{k}} \phi_{\mathbf{k}}(t) e^{i\mathbf{k}\cdot\mathbf{x}} + a_{\mathbf{k}}^\dagger \phi_{\mathbf{k}}^*(t) e^{-i\mathbf{k}\cdot\mathbf{x}} \right], \quad (37)$$

and define

$$\phi(x) = \chi(x) + \varphi(x) \quad (38)$$

where in $\chi(x)$ the integral over d^3k runs only over the comoving momenta with $|\mathbf{k}| < k_*(t)$ while in $\varphi(x)$ it runs only over comoving momenta with $|\mathbf{k}| > k_*(t)$. The value k_* that separates the two regimes can be defined by rewriting eq. (36) as $\omega_{k_*} = \nu H(t)$ where ν is a dimensionless constant with $\nu \gtrsim 1$.⁹ Therefore

$$k_*(t) = a(t) \sqrt{\nu^2 H^2(t) - m^2}. \quad (39)$$

⁸ Our approach here differs from that in [34], where the authors introduce an ultralight scalar field with $m \sim H_0$, and then derive an effective action for the gravitational field only, by integrating out all modes of the scalar field, including its low-frequency modes.

⁹ As usual in this separation the constant ν is in principle arbitrary, and the evolution of the low-energy effective theory with respect to ν is described by renormalization group equations.

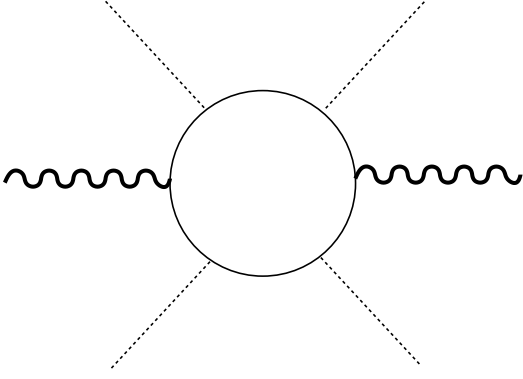


Figure 1: A Feynman diagram that contributes to the generation of a term $f(\chi)R$ in the effective action. The wavy line represents the classical gravitational field, the solid line in the loop is the field φ and the dashed line represents the field χ . A graph with n external χ lines contributes a term $\propto \chi^n R$ to the effective action (here $n = 4$).

Of course, the low-momentum modes exist only if m and t are such that $m \leq \nu H(t)$. At times where $m \gg H(t)$, all modes become heavy and ϕ can be integrated out completely, or equivalently $k_*(t) = 0$. Observe that the separation between χ and φ is time-dependent. Denoting all other massive matter fields collectively by ψ , the effective action now becomes a functional of both the metric and χ , given by

$$e^{iS_{\text{eff}}[g_{\mu\nu}, \chi]} = \int \mathcal{D}\psi \mathcal{D}\varphi e^{iS_m[g_{\mu\nu}, \chi + \varphi, \psi]}. \quad (40)$$

In principle $S_{\text{eff}}[g_{\mu\nu}, \chi]$ will include all possible terms that are consistent with general covariance.

In particular, there can be a term proportional to $\int d^4x \sqrt{-g} f(\chi)R$, where $f(\chi)$ is some function whose explicit form depends on the details of the original action. Diagrammatically, this term comes from graphs such as that in fig. 1, where the wavy lines represent the background gravitational field, the solid line circulating in the loop represents the high-frequency modes φ and the external dashed lines represent χ . A graph with the insertion of n external χ lines contributes a term proportional to $\chi^n R$ to the effective action. Observe that the χ field only occurs in external lines, since it is not integrated over. Further notice that a graph with two external wavy lines produces the term in R quadratic in the background metric. Terms cubic and of higher order in the background metric coming from R are obtained from graphs with three or more external wavy lines. The integral in the loop runs only over high-frequency momenta, i.e. $k_* < |k| < \infty$, and can be regularized in the UV using e.g. dimensional regularization.

The graphs shown in fig. 1 assumes the existence of a ϕ^3 interaction term in the potential $V(\phi)$. After splitting $\phi = \chi + \varphi$ this produces a vertex $\chi\varphi\varphi$. With a $\lambda\phi^4$ term in the potential we had vertices such as $\chi\chi\varphi\varphi$ instead,

where two external χ lines are attached at each vertex in the loop. Furthermore, the gravitational field must be similarly split as $g_{\mu\nu} = \bar{g}_{\mu\nu} + h_{\mu\nu}$, where the background $\bar{g}_{\mu\nu}$ contains the low frequency modes, and basically coincide with the background FRW metric, while $h_{\mu\nu}$ contains the high-frequency modes and describes quantum fluctuations over the background. In eq. (40) we must therefore integrate also over $h_{\mu\nu}$, so there are also graphs such as that in fig. 1 in which the internal φ loop is replaced by a loop where $h_{\mu\nu}$ circulates. The non-linear terms in the gravitational action provide vertices such as $\bar{g}hh$ that allow us to attach external $\bar{g}_{\mu\nu}$ lines to the $h_{\mu\nu}$ loop. Similarly the interaction of ϕ with gravity provides vertices such as χhh that allows us to attach external χ lines to the loop. Therefore terms of the form $f(\chi)R$ are generated through the gravitational interaction even if $V(\phi) = (1/2)m^2\phi^2$ and there is no direct interaction between χ and φ . All these graphs are generically non-zero, and therefore a term $f(\chi)R$ is in general generated in the effective action, even if there was no term $f(\phi)R$ in the original action for ϕ .

Among all terms that can be generated by the integration over the high-frequency modes, at the phenomenological level the term $f(\chi)R$ is the most interesting one, since it involves the lowest number of derivatives. Therefore it is not suppressed at scales below the Planck mass, contrary to terms involving R^2 , $R_{\mu\nu}R^{\mu\nu}$, etc. or interaction terms involving derivatives of χ . Therefore the most general effective action, for generating a cosmologically relevant $\langle 0|T_{\mu\nu}|0\rangle$, is of the form

$$S_{\text{eff}}[g_{\mu\nu}, \chi] = \int d^4x \sqrt{-g} \times \left[\frac{1 - f(\chi)}{16\pi G} R - \frac{1}{2} g^{\mu\nu} \partial_\mu \chi \partial_\nu \chi - V(\chi) \right], \quad (41)$$

plus higher-derivative operators, that are not relevant at scale much below the Planck scale, and possibly non-minimal kinetic terms, as in k-essence models.

The integration over the high-frequency modes in FRW has also been studied, in a somewhat different language, in the context of stochastic inflation [35–37]. For instance, for a minimally coupled scalar field ϕ in De Sitter space, the result for $S_{\text{eff}}[g_{\mu\nu}, \chi]$ can be written in the form [36]

$$e^{iS_{\text{eff}}[g_{\mu\nu}, \chi]} = \int \mathcal{D}\xi P[\xi] e^{iS_0[g_{\mu\nu}, \chi, \xi]} \quad (42)$$

where S_0 is an action from which one derives the classical equation of motion for χ , in the form

$$\ddot{\chi} + 3H\dot{\chi} - a^{-2}\nabla^2\chi + V'(\chi) = \dot{\xi} + 3H\xi. \quad (43)$$

Here ξ is a random field with a two-point correlator

$$\langle \xi(x)\xi(x') \rangle = A(x - x'), \quad (44)$$

and $P[\xi]$ in eq. (42) is given by

$$P[\xi] = \exp \left\{ -\frac{1}{2} \int d^4x d^4x' \xi(x) A^{-1}(x, x') \xi(x') \right\}. \quad (45)$$

Defining again the high-frequency modes as the modes with $\omega_k > \nu H$, the form of $A(x - x')$ in the most interesting limit $\nu \gtrsim 1$ has not been computed, while in the limit $\nu \ll 1$ it is given by [36]

$$A(x - x') = \frac{H^3}{4\pi^2} \frac{\sin(\nu H a r)}{\nu H a r} \delta(t - t') + \mathcal{O}(\nu^2), \quad (46)$$

where $r = |\mathbf{x} - \mathbf{x}'|$. The above formulation is physically quite intuitive as it allows us to interpret χ as a classical field driven by some stochastic noise. This is the typical picture of stochastic inflation. If one was to integrate out ξ from eq. (42) one would find the effective action $S_{\text{eff}}[g_{\mu\nu}, \chi]$ that can be directly obtained from diagrams such as that in fig. 1.

It is also instructive to directly compute $\langle 0|T_{\mu\nu}|0\rangle$ based on the sum of zero-point energies, and to see how a term $f(\chi)R$ could be generated in the effective action even if it was not present at first. In the usual setting with only massive fields, e.g. a massive scalar field Φ , one expands the field in terms of creation and annihilation operators

$$\Phi(x) = \int \frac{d^3k}{(2\pi)^3 \sqrt{2k}} \left[a_{\mathbf{k}} \Phi_k(t) e^{i\mathbf{k}\cdot\mathbf{x}} + a_{\mathbf{k}}^\dagger \Phi_k^*(t) e^{-i\mathbf{k}\cdot\mathbf{x}} \right]. \quad (47)$$

The modes $\Phi_k(t)$ are obtained by solving the equation of motion in the given FRW metric, here

$$\Phi_k'' + 2\frac{a'}{a}\Phi_k' + k^2\Phi_k = 0, \quad (48)$$

where the prime denotes derivatives with respect to conformal time η . Then the vacuum energy density is given by

$$\langle 0|T_{00}|0\rangle = \frac{1}{2} \int \frac{d^3k}{(2\pi)^3 2k} \left(|\dot{\Phi}_k|^2 + \frac{k^2}{a^2} |\Phi_k|^2 \right), \quad (49)$$

where some regularization of the momentum integral is understood, e.g. a cutoff Λ_c over momenta or dimensional regularization. Thus, according to the standard rules of semiclassical gravity, the field Φ is treated as a quantum field, while the metric $g_{\mu\nu}$ is treated classically, and the modes of the quantum field are determined solving an equation in this classical background. The energy density $\langle 0|T_{00}|0\rangle$ calculated in this way depends explicitly on the Hubble parameter $H(t)$ because the modes $\Phi_k(\eta)$ depend on the background metric, in this case on the scale factor $a(\eta)$ and conformal time $|\eta| \sim (aH)^{-1}$ of FRW. For example, in a De Sitter epoch or during matter dominance (MD),

$$\Phi_k(\eta) = \frac{1}{a(\eta)} \left(1 - \frac{i}{k\eta} \right) e^{-ik\eta}. \quad (50)$$

In the limit $k\eta \rightarrow \infty$, and therefore $k_{\text{phys}} = k/a \gg H$, these modes reduce to the usual plane waves of flat space (this corresponds to the most natural choice of vacuum state, namely the Bunch-Davies vacuum; the most general form of the modes is a superposition of positive

and negative frequency modes with arbitrary Bogoliubov coefficients). Therefore the leading UV divergence in eq. (49) is the same as in flat space, i.e. the Λ_c^4 term if we regularize with a cutoff in momentum space. However, the term $i/k\eta$ in eq. (50), as well as the time dependence in the factor $1/a(\eta)$, affect the subleading divergence, and produce a quadratic divergence $\propto H^2(t)\Lambda_c^2$. (For details on these calculations see e.g. ref. [10].)

When there is also an ultra-light scalar field ϕ the above scenario changes, because ϕ must be split into its low-frequency modes χ and its high-frequency modes φ . The field χ should be treated as classical on the same footing as the metric $g_{\mu\nu}$ since, on physical grounds, modes with wavelength larger than the horizon are expected to undergo decoherence. (See ref. [37] for a more careful discussion of this point.) The high-frequency modes φ should instead be quantized in the classical background, given both by $g_{\mu\nu}$ and χ . In particular, the equation of motion for φ_k that replaces eq. (48) will depend both on the scale factor and on the homogeneous field $\chi(\eta)$. It is derived from the equation of motion of ϕ by setting $\phi = \chi + \varphi$ and expanding φ in momentum modes φ_k . As a result the modes (50) will acquire a dependence on χ which will generically translate into a χ -dependence of $\langle 0|T_{\mu\nu}|0\rangle$. In general, this produces all terms which are consistent with the symmetries of the problem, and therefore also terms corresponding to $f(\chi)R$ in the effective action.

Finally, we observe that similar considerations hold for massless conformally-coupled fields, including the electromagnetic field. If it were not for the conformal anomaly, in FRW (which is conformally equivalent to Minkowski space) these fields would decouple from the metric and could be integrated out trivially. Because of the conformal anomaly at the one-loop level these fields do not decouple but still they can be integrated out exactly, giving rise to the anomaly-induced effective action. However the fact that, technically, it is possible to integrate them out exactly does not mean that, physically, this is the correct thing to do. In any case, the long-wavelength modes of these fields should be kept in the effective action, on the same footing as the classical metric, and produce a classical background (whether of a scalar field, of an electromagnetic field, etc.) which in general will be relevant on cosmological scales. Excluding from the integration the modes with physical momenta $k_{\text{phys}} < k_*/a(t) \sim H(t)$ could also in principle affect the non-local terms of the anomaly-induced effective action.

B. Interacting vacuum fluctuations

We next discuss the properties of $\langle 0|T_{\mu\nu}|0\rangle$ derived from the effective action (41). Models of this form have been studied in the literature as generalizations of the simplest quintessence models [38–43]. Of course such models can be introduced and investigated in their own right, without being related to zero-point fluctuations.

Our line of arguments leading to such models, based on the most general form of $\langle 0|T_{\mu\nu}|0\rangle$ in the presence of ultra-light scalar fields, gives further motivation for their study. This insight will also allow us the recast the cosmological equations of the model in an interesting form, which can be seen to describe the interaction between vacuum fluctuation of the field ϕ and the classical part χ of the same field.

Specializing to a FRW metric and to a spatially homogeneous field $\chi = \chi(t)$, the equations of motion derived from the action (41) are [38–43]

$$H^2 = \frac{8\pi G}{3}(\rho_B + \rho_\chi), \quad (51)$$

$$2\dot{H} + 3H^2 = -8\pi G(p_B + p_\chi), \quad (52)$$

$$\ddot{\chi} + 3H\dot{\chi} + V'(\chi) = -\frac{3}{8\pi G}(\dot{H} + 2H^2)f'(\chi), \quad (53)$$

where ρ_B , p_B are the energy density and pressure of the matter-radiation background, respectively, and ρ_χ , p_χ are those associated to χ as given below. The prime denotes the derivative with respect to χ . Denoting the total EOS parameter by $w_{\text{tot}} \equiv (p_B + p_\chi)/(\rho_B + \rho_\chi)$, the two Friedmann equations are combined to find

$$\dot{H} = -\frac{3}{2}H^2(1 + w_{\text{tot}}), \quad (54)$$

such that eq. (53) can be recast as

$$\ddot{\chi} + 3H\dot{\chi} + V'(\chi) = -\frac{3}{16\pi G}H^2(1 - 3w_{\text{tot}})f'(\chi). \quad (55)$$

It is convenient to separate ρ_χ and p_χ into two parts,

$$\rho_\chi = \rho_X + \rho_Z, \quad p_\chi = p_X + p_Z, \quad (56)$$

where

$$\rho_X = \frac{1}{2}\dot{\chi}^2 + V(\chi), \quad p_X = \frac{1}{2}\dot{\chi}^2 - V(\chi), \quad (57)$$

are the standard contributions of the classical part, while

$$\rho_Z = \frac{3}{8\pi G} \left[H^2 f(\chi) + H\dot{f}(\chi) \right], \quad (58)$$

$$p_Z = -\frac{1}{8\pi G} \left[(2\dot{H} + 3H^2)f(\chi) + 2H\dot{f}(\chi) + \ddot{f}(\chi) \right]. \quad (59)$$

Using again eq. (54), eq. (59) can also be rewritten as

$$p_Z = \frac{1}{8\pi G} \left[3w_{\text{tot}}H^2 f(\chi) - 2H\dot{f}(\chi) - \ddot{f}(\chi) \right]. \quad (60)$$

The subscript Z , which stands for “zero-point fluctuations”, stresses that in our approach ρ_Z and p_Z can be interpreted as a contribution coming from vacuum fluctuations, i.e. from the VEV $\langle 0|T_{\mu\nu}|0\rangle$ obtained through eq. (1) from an effective action that includes the contribution $f(\chi)R$ which generically appears in the presence of the ultra-light field χ .

A rather interesting property of ρ_Z is that it is of the order of the critical density at all times. For instance,

recalling that the critical density at time t is $\rho_c(t) = (3/8\pi G)H^2(t)$, we see that the term $(3/8\pi G)H^2(t)f(\chi)$ in ρ_Z is just $f(\chi)\rho_c(t)$. In the slow-roll regime $\chi(t)$ is roughly constant, and this term provides an approximately fixed fraction of the critical energy density. In this way we obtain a term of order $M_{\text{Pl}}^2 H^2(t)$ in the energy density, which is not reabsorbed into a renormalization of Newton’s constant because its prefactor contains the function $f(\chi)$.

The above equations can be rewritten in a form which allows for an appealing physical interpretation, namely that the “classical” energy-momentum tensor described by ρ_X and p_X interacts with the vacuum fluctuations, described by ρ_Z and p_Z . The interaction is governed by a coupled energy-momentum conservation equation (thereby providing an explicit realization of a mechanism proposed in [10, 11]). To this purpose, it is convenient to transform the time derivatives into derivatives with respect to $x \equiv \ln a$, which we denote by ∂_x . By means of eq. (54) we can then write

$$\rho_Z = \rho_c(f + \partial_x f), \quad (61)$$

$$p_Z = w_Z \rho_Z, \quad (62)$$

$$w_Z = w_{\text{tot}} - \frac{(1 + 3w_{\text{tot}})\partial_x f + 2\partial_x^2 f}{6(f + \partial_x f)}. \quad (63)$$

If the field χ does not exchange energy with the other forms of matter included in the background action S_B (as we have indeed already assumed when deriving eq. (53)), its energy density and pressure satisfy the conservation equation

$$\dot{\rho}_\chi + 3H(\rho_\chi + p_\chi) = 0. \quad (64)$$

Splitting ρ_χ and p_χ as in eq. (56) and defining

$$w_X \equiv \frac{p_X}{\rho_X} = \frac{(1/2)\dot{\chi}^2 - V(\chi)}{(1/2)\dot{\chi}^2 + V(\chi)}, \quad (65)$$

we can write the continuity equation as

$$\dot{\rho}_Z + 3H(1 + w_Z)\rho_Z = Q, \quad (66)$$

$$\dot{p}_X + 3H(1 + w_X)\rho_X = -Q, \quad (67)$$

where $Q(t)$ is the energy density transfer rate from X to Z . Inserting the explicit expressions for ρ_Z and w_Z given in eqs. (61) and (63) we find

$$Q = \frac{1}{2}(1 - 3w_{\text{tot}})H(t)\rho_c(t)\partial_x f. \quad (68)$$

Since $\rho_c(t)$ and $H(t)$ provide the natural scales for the energy density and for the inverse of time, respectively, a natural scale for the rate of energy transfer is $H(t)\rho_c(t)$. We see from eq. (68) that, with respect to this scale, the actual rate Q differs by a factor proportional to $\partial_x f$ and to $(1 - 3w_{\text{tot}})$. Since $H\partial_x f = \dot{f}$, we see that the actual rate of energy transfer is governed by the time variation of $f(x)$, and therefore in a slow-roll phase $|Q| \ll H(t)\rho_c(t)$.

Furthermore, during RD $w_{\text{tot}} \approx 1/3$, providing a further suppression factor. However, note that Q does in general not vanish exactly in RD because of the contributions from p_Z and matter. We will see this for a typical example in the next section.

Conservation equations with energy exchange such as eqs. (66) and (67), that even go back to Bronstein (1933) (see [5]), are at the basis of much recent work on interacting dark energy/matter models (see e.g. [39, 44–49] and references therein). Typically the functional form of $Q(t)$ is chosen on purely phenomenological grounds. It is interesting to see that in this model the form of $Q(t)$ is explicitly predicted in terms of $\partial_x f$.¹⁰

In summary, the energy-momentum tensor of the classical part, $T_{\mu\nu}^0 = (\rho_X, a^2 \delta_{ij} p_X)$ and that of the vacuum fluctuations $\langle 0|T_{\mu\nu}|0\rangle = (\rho_Z, a^2 \delta_{ij} p_Z)$ are in general not separately conserved:

$$\nabla^\mu \langle 0|T_{\mu\nu}|0\rangle = -\nabla^\mu T_{\mu\nu}^0 \neq 0. \quad (69)$$

Rather, they interact through a coupled energy-momentum conservation equation.

C. Cosmological evolution and tracking dark energy

The cosmological evolution obtained from eqs. (51)–(53) has been studied in detail, see e.g. refs. [38–43]. To illustrate the typical evolution, we consider a model with

$$\frac{f(\chi)}{16\pi G} \equiv F(\chi) = \frac{1}{2}\xi\chi^2, \quad V(\chi) = \frac{1}{2}m^2\chi^2, \quad (70)$$

and we take for definiteness $\xi > 0$. It is useful to introduce the dimensionless parameter

$$\mu \equiv \frac{m}{\xi^{1/2}H_0}. \quad (71)$$

The ratio of the term $6H^2 F$ in ρ_Z to the potential V is equal to $6H^2/(\mu^2 H_0^2)$. Therefore, if $\mu = 3$, these two terms become comparable at the present epoch, while in the past $6H^2 F \gg V$. For larger values of μ the potential V begins to give a significant contribution to the energy density somewhat earlier than today, while for smaller μ it will only become important in the future. The evolution of χ , determined by eq. (55), is different depending on whether $V'(\chi)$ or $3H^2(1-3w_{\text{tot}})F'(\chi)$ dominates. The ratio of these two terms is

$$\frac{3H^2(1-3w_{\text{tot}})F'(\chi)}{V'(\chi)} = \frac{3H^2(1-3w_{\text{tot}})}{\mu^2 H_0^2}. \quad (72)$$

¹⁰ To make the connection with the discussion just after eq. (15) it is useful to introduce the EOS parameter $w_*(t)$ defined by the condition $\dot{\rho}_Z + 3H(1+w_*)\rho_Z \equiv 0$. Using eqs. (61) and (54) we get $w_* = w_{\text{tot}} - (\partial_x f + \partial_x^2 f)/[3(f + \partial_x f)]$. Observe that $\partial_x f = 0$ implies $w_* = w_{\text{tot}}$ consistent with eq. (15). Comparing with eq. (63) we see that $w_Z = w_* + \bar{w}$ where $\bar{w} = [1-3w_{\text{tot}}]\partial_x f/[6(f + \partial_x f)]$ and therefore in general $w_Z \neq w_*$.

Since observationally $w_{\text{tot}}^0 \approx -0.7$ today, this ratio becomes of order unity at the present epoch for $\mu \simeq 3$. In the past it was much larger than one. This is true even in RD despite the fact that, during RD, $(1-3w_{\text{tot}}) \ll 1$. To see this in a simpler setting, let us at first include just the contribution of matter to w_{tot} in the RD phase. Including both matter and radiation, deep in RD where $\rho_M \ll \rho_R$ and the scale factor $a \ll a_{\text{eq}}$ (where $a_{\text{eq}} \equiv \Omega_R^0/\Omega_M^0 \approx 3 \times 10^{-4}$) we have

$$w_{\text{tot}} = \frac{(1/3)\rho_R}{\rho_R + \rho_M} = \frac{1}{3} \frac{1}{1 + a/a_{\text{eq}}}. \quad (73)$$

Therefore

$$(1-3w_{\text{tot}})H^2 \simeq \frac{a}{a_{\text{eq}}} H^2. \quad (74)$$

In RD $a \propto t^{1/2}$ and $H^2 \propto 1/t^2$, so $(1-3w_{\text{tot}})H^2 \propto t^{-3/2}$ and, going backward in time, the growth of H^2 more than compensate the decrease of $(1-3w_{\text{tot}})$.

More importantly, in this model there is a significant contribution from DE to w_{tot} even in RD. This contribution, contrary to that of matter, is not parametrically suppressed in the limit $a/a_{\text{eq}} \ll 1$. To compute it in the matter and radiation era we can safely neglect p_X . Then we find from the expressions in the last section

$$1-3w_{\text{tot}} \simeq \frac{(1-f-\partial_x f)/(1+a_{\text{eq}}/a) + \partial_x f + \partial_x^2 f}{1-f-\partial_x f/2}, \quad (75)$$

which correctly reduces to eq. (73) when $f = \text{const}$. Thus, the combination $(1-3w_{\text{tot}})$ is generically non-zero, and for $\mu = \mathcal{O}(1)$ the evolution of χ during RD and most of MD is dominated by the non-minimal coupling $F(\chi)R$. The potential $V(\chi)$ only becomes relevant when approaching the current epoch, while for larger values of μ the potential becomes relevant earlier. Eventually, the field χ will exit the slow-roll phase and perform damped oscillations around the minimum of $V(\chi)$ at $\chi = 0$.

In this model the total DE density $\rho_X + \rho_Z$ tracks the total energy density $\rho_c(t)$, as we can see from fig. 2. This is basically due to the combination of two factors: during RD and deep into MD the tracking is ensured by the term $f(\chi)\rho_c(t)$ in ρ_Z . This term, however, cannot be the dominant one today, since it is proportional to $H^2(t)$, and in the recent cosmological epoch a time dependence $\rho_{\text{DE}} \propto H^2(t)$ is observationally excluded. Thus, approaching the recent epoch, we must switch to a regime where ρ_X dominates and χ is still approximately slowly-rolling such that w_X is close to -1 , see eq. (65), and the DE density is approximately constant.

With suitable values of the parameters ξ , μ and of the initial conditions we can get values for Ω_{DE} and w_{DE} today of the order of the observationally preferred ones. For instance, in fig. 2 we used $\mu = 24.3$, $\xi = 5 \times 10^{-4}$ and we set the initial conditions $\chi_{\text{in}} = 4$ and $\chi'_{\text{in}} = 0$ deep in RD (measuring χ in units $8\pi G = 1$). With these values we get $\Omega_{\text{DE}} \simeq 0.73$ and $w_{\text{DE}} = -0.952$ today.

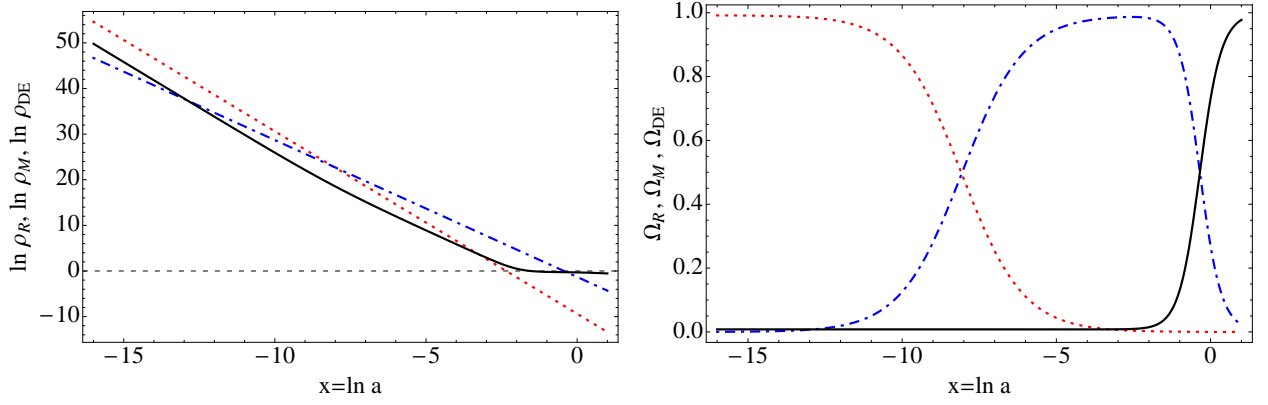


Figure 2: *Left*: $\log \rho$ for radiation (red, dotted), matter (blue, dot-dashed) and dark energy (black solid) as functions of $\log a$. *Right*: the same plot on a linear vertical scale, in terms of $\Omega_i = \rho_i(t)/\rho_c(t)$, with $i = R, M, DE$. We used the values $\mu = 24.3$, $\xi = 5 \times 10^{-4}$ and the initial conditions $\chi_{\text{in}} = 4$ (in units $8\pi G = 1$), $\chi'_{\text{in}} = 0$ deep in RD.

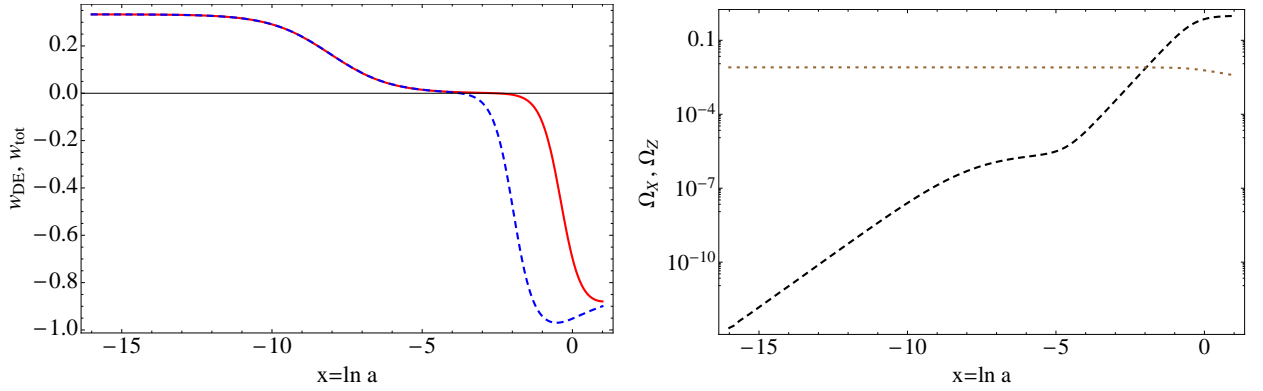


Figure 3: *Left*: w_{tot} (red, solid) and w_{DE} (blue, dashed) as functions of $\log a$ for the same parameters as in fig. 2. *Right*: the separate contributions $\Omega_X = \rho_X(t)/\rho_{\text{tot}}(t)$ (black, dashed) and $\Omega_Z = \rho_Z(t)/\rho_{\text{tot}}(t)$ (brown, dotted).

The evolution of the EOS parameter $w_{DE} \equiv p_{DE}/\rho_{DE}$ is shown in the left panel of fig. 3, together with the evolution of w_{tot} . This shows that a realistic cosmological evolution can be obtained in this model.¹¹

It should however be stressed that the coincidence problem is not really solved by the fact that the DE density tracks the total energy. The total dark energy density $\rho_{DE} = \rho_X + \rho_Z$ switches from an early regime where it is dominated by ρ_Z , and hence evolves approximately as $H^2(t)$, to a late regime where it is dominated by ρ_X , and therefore is approximately constant. The transition between these regimes can nicely be observed when looking at the separate evolution of $\rho_X(t)/\rho_c(t)$ and $\rho_Z(t)/\rho_c(t)$, plotted in the right panel of fig. 3 for

the same model parameters used for the other figures. The time when the transition takes place is controlled by ratio between the mass of χ and the Hubble constant. In fact, it is the parameter $\mu = m/(\xi^{1/2}H_0)$ that is crucial here: the larger μ the earlier the transition. As we argued above, if $\mu = 3$ the transition happens about at the current epoch. Notice though, compared to Λ CDM, the coincidence problem is still sensibly alleviated. This is clearly seen from fig. 4, where the ratio $\rho_{DE}(t)/\rho_{\text{tot}}(t)$, computed in the model with $F(\chi)$ and $V(\chi)$ given in eq. (70), is compared to the same ratio in Λ CDM.

Exploring the parameter space to a certain extent, we find that the figures shown represent the typical behavior of this model, as long reasonable initial conditions are set. In principle, a deeper study of the stability and the parameter space of the system would be needed, but it is not the aim of this work to do so. The reason is the lack of motivation for the specific choices for $F(\chi)$ and $V(\chi)$ at this point. Therefore we will discuss a more general if slightly approximate approach in the following.

¹¹ Furthermore, in this model the limits on the time variation of Newton's constant require $32\pi G(\xi\chi_0)^2 < \omega_{\text{BD}}^{-1}$, where χ_0 is the present value of $\chi(t)$ and ω_{BD} is the Brans-Dicke parameter [41]. Using the bound $\omega_{\text{BD}} > 4.2 \times 10^4$ inferred by the Cassini mission [50] and measuring χ in units $8\pi G = 1$, this bound reads $\xi\chi_0 < 2.4 \times 10^{-3}$. This is satisfied by the solution shown in fig. 2, for which $\xi\chi_0 \approx 1.9 \times 10^{-3}$.

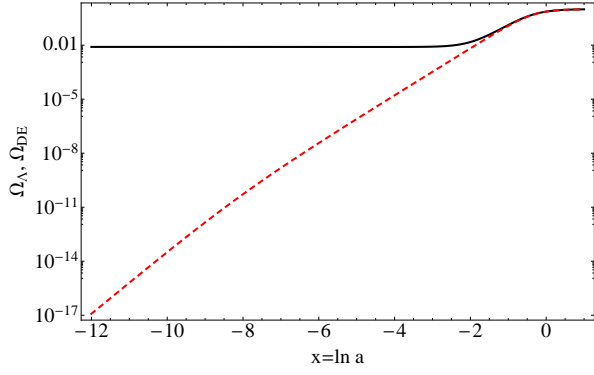


Figure 4: The ratio $\Omega_{\text{DE}}(t) \equiv \rho_{\text{DE}}(t)/\rho_{\text{tot}}(t)$, computed in the model with $F(\chi)$ and $V(\chi)$ given in eq. (70) (solid, black) compared to the ratio $\Omega_{\Lambda}(t) \equiv \rho_{\Lambda}/\rho_{\text{tot}}(t)$ in Λ CDM (dashed, red).

D. Comparison with observations

The model that we have discussed has a DE density $\rho_{\text{DE}}(t) = \rho_X(t) + \rho_Z(t)$, where $\rho_X(t)$ is the typical quintessence energy density due to the rolling of the scalar field χ in the potential $V(\chi)$, while $\rho_Z(t)$ comes from the term $F(\chi)R$ in the effective action. With a suitable choice of the mass m in $V(\chi)$ the term ρ_X comes to dominate near the present epoch, while it is totally negligible at early times, and its EOS parameter w_X is close to -1 today, as required by the observations. In contrast, $\rho_Z(t)$ is proportional to $H^2(t)$. In order to have a viable model, ρ_Z must be subleading today since a DE term proportional to $H^2(t)$ is observationally excluded (see e.g. [47–49]). However, the fact that this term is approximately a constant fraction of $\rho_c(t)$ means that its integrated effect over the whole history of the Universe could give observable effects.

1. A more general parametrization of ρ_{DE}

The aim of this section is to perform a comparison of such a model with cosmological observations, expanding on the results that we have presented in [11]. A possible way to proceed could be to choose a particular form of the functions $F(\chi)$ and $V(\chi)$, such as those given in eq. (70), and find the corresponding limits on the parameters ξ and m . However, such limits would be tied to the specific choice made for the functions $F(\chi)$ and $V(\chi)$. In the absence of strong theoretical motivations for any such choice, it seems more useful to extract some general features of the DE sector of such models, and parametrize the DE density in a form that might have a more general validity. This can be done by recasting eqs. (57) and (58) in the form

$$\rho_{\text{DE}} = V + H^2 \left[\frac{1}{2} (\partial_x \chi)^2 + 6(F + \partial_x F) \right], \quad (76)$$

where we have again transformed the derivative with respect to t into the derivative with respect to $x \equiv \ln a$, denoted by ∂_x . By means of the Friedmann equation, $H^2 = (8\pi G/3)(\rho_R + \rho_M + \rho_{\text{DE}})$, and the definition

$$Z \equiv \frac{8\pi G}{3} \left[\frac{1}{2} (\partial_x \chi)^2 + 6(F + \partial_x F) \right] \quad (77)$$

we can write the DE density in the form

$$\rho_{\text{DE}} = \frac{V + (\rho_R + \rho_M)Z}{1 - Z}. \quad (78)$$

Finally, we implicitly define the three functions $w_0(a)$, $\bar{\epsilon}_R(a)$ and $\bar{\epsilon}_M(a)$ and the constant C by writing

$$\rho_{\text{DE}} = Ca^{-3(1+w_0(a))} + \rho_R(a)\bar{\epsilon}_R(a) + \rho_M(a)\bar{\epsilon}_M(a), \quad (79)$$

such that

$$\bar{\epsilon}_R(a) = \bar{\epsilon}_M(a) = \epsilon(a) \equiv \frac{Z}{1 - Z}, \quad (80)$$

$$w_0(a) = -1 - \frac{1}{3} \ln \left[\frac{V}{CZ} \epsilon(a) \right]. \quad (81)$$

When w_0 is constant, the term $Ca^{-3(1+w_0)}$ in eq. (79) corresponds to the usual dark energy density of a minimal quintessence model. On the other hand, the presence of the terms $\rho_R\bar{\epsilon}_R$ and $\rho_M\bar{\epsilon}_M$ clearly displays the tracking behavior of the DE density. Choosing a specific model for the functions $F(\chi)$ and $V(\chi)$ corresponds to making a choice for $w_0(a)$ and $\epsilon(a)$. The constant C is fixed by the DE density today, or better, by a combination of the model parameters in F and V and the initial value of χ .

Going back to the example discussed in the last section, $F(\chi) = (1/2)\xi\chi^2$ and $V(\chi) = (1/2)m^2\chi^2$, we solve eqs. (51) and (53) numerically and show the evolution of the functions $w_0(a)$ and $\epsilon(a)$ in fig. 5. We see that in the recent epoch where the term $Ca^{-3(1+w_0)}$ is relevant w_0 can roughly be approximated by a constant, within a few percent. Moreover, the function $\epsilon(a)$ is constant to excellent approximation in RD and decreases linearly with $\ln a$ in MD. This is the reason why we separated the terms into $\rho_R\bar{\epsilon}_R$ and $\rho_M\bar{\epsilon}_M$, as either one will be negligible when the other one is dominating. Although $\epsilon(a)$ raises again when the present epoch is approached, see fig. 5, at this stage the contribution from $\rho_M\bar{\epsilon}_M$ in the DE density is subdominant compared to the term $Ca^{-3(1+w_0)}$ coming from ρ_X (see the right panel in fig. 3). Accordingly, the behavior of $\epsilon(x)$ in this regime is not very relevant. Therefore, we can approximately write

$$\bar{\epsilon}_R(a) \simeq \epsilon_R \quad (82)$$

$$\bar{\epsilon}_M(a) \simeq \epsilon_M + B \ln a \quad (83)$$

with independent constant parameters ϵ_R and ϵ_M . The constant B , on the other hand, is fixed by requiring the two branches to coincide at equality: $B = (\epsilon_R - \epsilon_M)/\ln a_{\text{eq}}$.

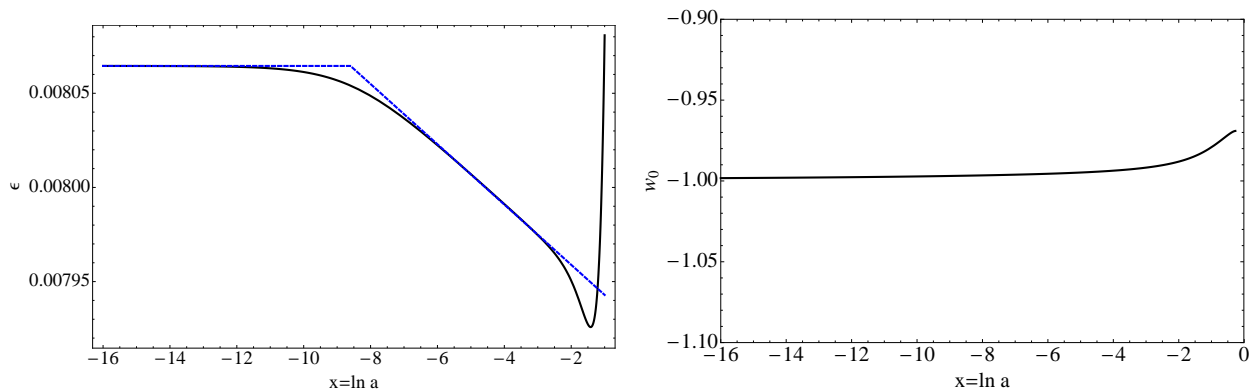


Figure 5: The evolution of the functions $\epsilon = Z/[1 - Z]$ and w_0 defined in eq. (79) are shown for the model and parameters of Sec. III C, see fig. 2. In the left panel we also show the $\epsilon_R - \epsilon_M$ approximation with $\epsilon_R = 0.00806$ and $\epsilon_M = 0.00793$ (blue, dashed).

The fact that $\bar{\epsilon}_R$ can be taken as constant is a consequence of χ rolling very slowly during RD, in fact it does essentially not move at all, see also ref. [43]. Thus, the function $Z(\chi)$ is virtually constant. The fact that in the MD era ϵ evolves linearly with $\ln a$ is because χ begins to roll, still slowly though, and therefore with $\ln a$ as this is the natural variable for cosmological evolution. Thus, this behavior appears to be quite general for typical viable models, and is likely to be not restricted to the specific model $F(\chi) = (1/2)\xi\chi^2$ and $V(\chi) = (1/2)m^2\chi^2$ that we have studied here. We will therefore use eq. (79) as our basic expression for the DE density, with $w_0(a) = w_0$ and $\epsilon(a)$ given in terms of the constants ϵ_R and ϵ_M . This model then is a generalization of the one studied in ref. [11], which corresponds to the limit $\epsilon_R = \epsilon_M$. It is a two-parameter generalization of w CDM, which in turn corresponds to the limit $\epsilon_R = \epsilon_M = 0$. We will generically use the label w ZCDM for this generalization (including the limiting case $\epsilon_R = \epsilon_M$ studied in [11]).

2. Results

We estimate the model parameters with a full likelihood analysis using modified versions of the publicly available Markov-Chain Monte-Carlo sampling code CosmoMC [51, 52] and the CMB-Boltzmann code CAMB [53, 54]. We assume a flat FRW background and put a prior on the age of the Universe to be between 10 and 20 Gyrs. Then we use the following datasets: constraints on the current Hubble parameter from Hubble Space Telescope (HST) observations [55] and on the number of relativistic degrees of freedom at Big-Bang Nucleosynthesis (BBN) [56]; the Union2 Compilation of Type Ia supernovae (SNe) of the Supernova Cosmology Project [57]; the angular power spectra of temperature and polarization anisotropies in the cosmic microwave background (CMB) from WMAP7 [58], ACBAR [59] and CBI [60]; and the Baryon Acoustic Oscillations (BAO) data from the Sloan Digital Sky Survey Data Release 7 (SDSS DR7)

[61].

We perform the analysis for the four following models: in w ZCDM2 we vary $\{w_0, \epsilon_R, \epsilon_M\}$; in w ZCDM1 we vary $\{w_0, \epsilon\}$ and set $\epsilon_R = \epsilon_M = \epsilon$; in w CDM we fix $\epsilon = 0$ in addition and, finally, in Λ CDM we fix $w_0 = -1$ as well. On top of these DE parameters we vary the common set of six base parameters. These are the slope n_s and amplitude $\log[10^{10}A_s]$ of the spectrum of primordial scalar curvature perturbations (modeled as a power law normalized at $k = 0.05 \text{ Mpc}^{-1}$), the depth to re-ionization τ , the physical baryon energy fraction $\Omega_b h^2$, the physical CDM energy fraction $\Omega_{\text{DM}} h^2$, and 100 times the ratio of the sound horizon to the angular diameter distance to the last-scattering surface, θ . We use flat priors for all parameters, set adiabatic initial conditions for the evolution of the cosmological perturbations and ignore vector and tensor modes for simplicity.

As opposed to the cosmological constant Λ , in the case of dynamical DE models we need to model the evolution of its inhomogeneities. If we were interested in a specific model, we should solve the perturbation equations derived from its Lagrangian. However, as discussed in Section III D 1, in the context of this work we are not really interested in analyzing a model with a specific form of the functions $F(\chi)$ and $V(\chi)$. We rather wish to extract more general features of a class of DE models suggested by our theoretical considerations. In this spirit, we will at first assume the DE to be a perfect fluid with a constant rest frame sound speed $\hat{c}_s = 1$, as it is usually done in the w CDM model, and we will later check how our results depend on this assumption.

The results are summarized in table I. We show the marginalized posterior probability distributions for w ZCDM1 and w ZCDM2 in fig. 6, comparing both cases with w CDM and Λ CDM. For w ZCDM1 we find lower and upper limits of

$$-1.25 < w_0 < -0.908, \quad -0.0201 < \epsilon < 0.0460, \quad (84)$$

at 95% C.L. For w ZCDM2 we find lower and upper limits

	Λ CDM		w CDM		w ZCDM1		w ZCDM2	
	$\langle \dots \rangle$	σ	$\langle \dots \rangle$	σ	$\langle \dots \rangle$	σ	$\langle \dots \rangle$	σ
n_s	0.966	0.011	0.963	0.013	0.967	0.014	0.971	0.016
$\log[10^{10} A_s]$	3.084	0.032	3.088	0.032	3.087	0.032	3.084	0.034
τ	0.087	0.014	0.086	0.014	0.087	0.015	0.089	0.015
$\Omega_b h^2$	0.023	0.001	0.023	0.001	0.0226	0.001	0.0225	0.001
$\Omega_{\text{DM}} h^2$	0.113	0.003	0.115	0.004	0.115	0.005	0.111	0.007
θ	1.040	0.002	1.040	0.002	1.039	0.003	1.042	0.006
w_0	-	-	-1.043	0.079	-1.067	0.087	-1.11	0.11
ϵ	-	-	-	-	0.0106	0.017	-	-
ϵ_M	-	-	-	-	-	-	0.041	0.048
ϵ_R	-	-	-	-	-	-	-0.0068	0.026

Table I: Means and standard deviations of the marginalized likelihoods. We observe that all additional DE parameters are consistent with Λ CDM at the 0.68% C.L. for all models w CDM, w ZCDM1 as well as w ZCDM2. All models have unit rest frame sound speed, $\hat{c}_s = 1$.

of

$$\begin{aligned}
 -1.35 < w_0 < -0.903, \quad & -0.0465 < \epsilon_M < 0.140, \\
 & -0.0558 < \epsilon_R < 0.0446,
 \end{aligned} \tag{85}$$

again at 95% C.L. In general, we observe a slight preference for models with $w_0 < -1$ and $\epsilon_M > 0$, i.e. early DE models with a recent phantom-crossing. However, all new parameters are consistent with their Λ CDM values at the 95% C.L. as this is also the case for w_0 in w CDM. The results are consistent with those for other early DE models found in the literature, e.g. [62–64]. However, note that our model has a slightly different evolution during radiation domination than the widely used parameterization proposed in ref. [65], and usually for that model phantom-crossing is excluded with the hard prior $w_0 \geq -1$ which we do not employ here.

In fig. 7 we show the combined likelihood contours for a subset of the parameters of w ZCDM1 and compare the constraints with those for the w CDM model. The slight degeneracy in the (ϵ, w_0) plane comes from the fact that $w_0(a)$ depends logarithmically on $\epsilon(a)$ as can be seen from eq. (80). For positive ϵ the data prefers $w_0 < -1$. Secondly, we note that the constraints on Ω_{DE} are virtually unchanged by introducing the new parameter ϵ , as there is basically no degeneracy in the $(\epsilon, \Omega_{\text{DE}})$ plane and the marginalized posterior of for Ω_{DE} is unchanged as well. Finally, ϵ is mildly correlated with the spectral index, n_s , as our analysis shows. This leads to a slight shift of the posterior to larger values of n_s in w ZCDM1 as compared to w CDM.

In fig. 8 we show the combined likelihood contours for w ZCDM2, comparing to w CDM. The correlation between ϵ_M and w_0 persists, which means that ϵ in the simplified w ZCDM1 model is more or less reflecting ϵ_M . This is because the observations of the SNe and BAO are more sensitive to the late-time than the early Universe expansion history. Introducing the additional parameter ϵ_R weakens the constraints on most parameters consid-

erably. This is also a consequence of the CMB data not being strongly sensitive to $\epsilon(a)$. Therefore, other high-redshift probes such as 21 cm line spectra, gamma-ray bursts and Lyman-alpha forest data need to be added to improve the constraints for such models considerably.

We next examine the dependence of our results on the form of the perturbations. It should be remarked that in Extended Quintessence (EQ) models, such as those represented by our 2-epsilons parametrization, the non-minimal coupling of the scalar field with the curvature leads to a specific behaviour of the perturbations, which is in general not equivalent to the one of a perfect fluid with unit rest-frame sound speed $\hat{c}_s = 1$. This issue was discussed in great details in refs. [66] and [67]. In [66] it was found that, since the fluctuation in the energy density of the scalar field, $\delta\rho_\chi$, is directly related to the perturbation in the Ricci scalar, $\delta\rho_\chi$ tracks the density perturbation of the dominant component of the Universe. In MD, in particular, $\delta_\chi \sim \delta_M$ on all sub-horizon scales, (where the density contrast is defined by $\delta_i = \delta\rho_i/\bar{\rho}_i$) leading to the possibility of formation of dark energy clumps. In others words, in EQ \hat{c}_s can drop much below unity, in contrast with what happens in normal quintessence where $\hat{c}_s = 1$ such that dark energy perturbations are strongly suppressed on all sub-horizon scales. In principle, this can affect the growth of perturbations in the matter sector and give observable effects. In ref. [66], it was determined that values such as $\hat{c}_s^2 \sim 0.001$ during MD are typical for EQ models. Moreover a small level of anisotropic stress is also present in EQ giving viscosity effects. In principle, the evolution of the perturbations in a given EQ model should be treated by solving the full system of perturbations equations for the specific choice of Lagrangian. Nevertheless, given the appropriate behaviour of the sound speed and the viscosity, the system can correctly be described in terms of fluid variables. To see how the constraints on our models depend on the sound speed we carry out the analysis

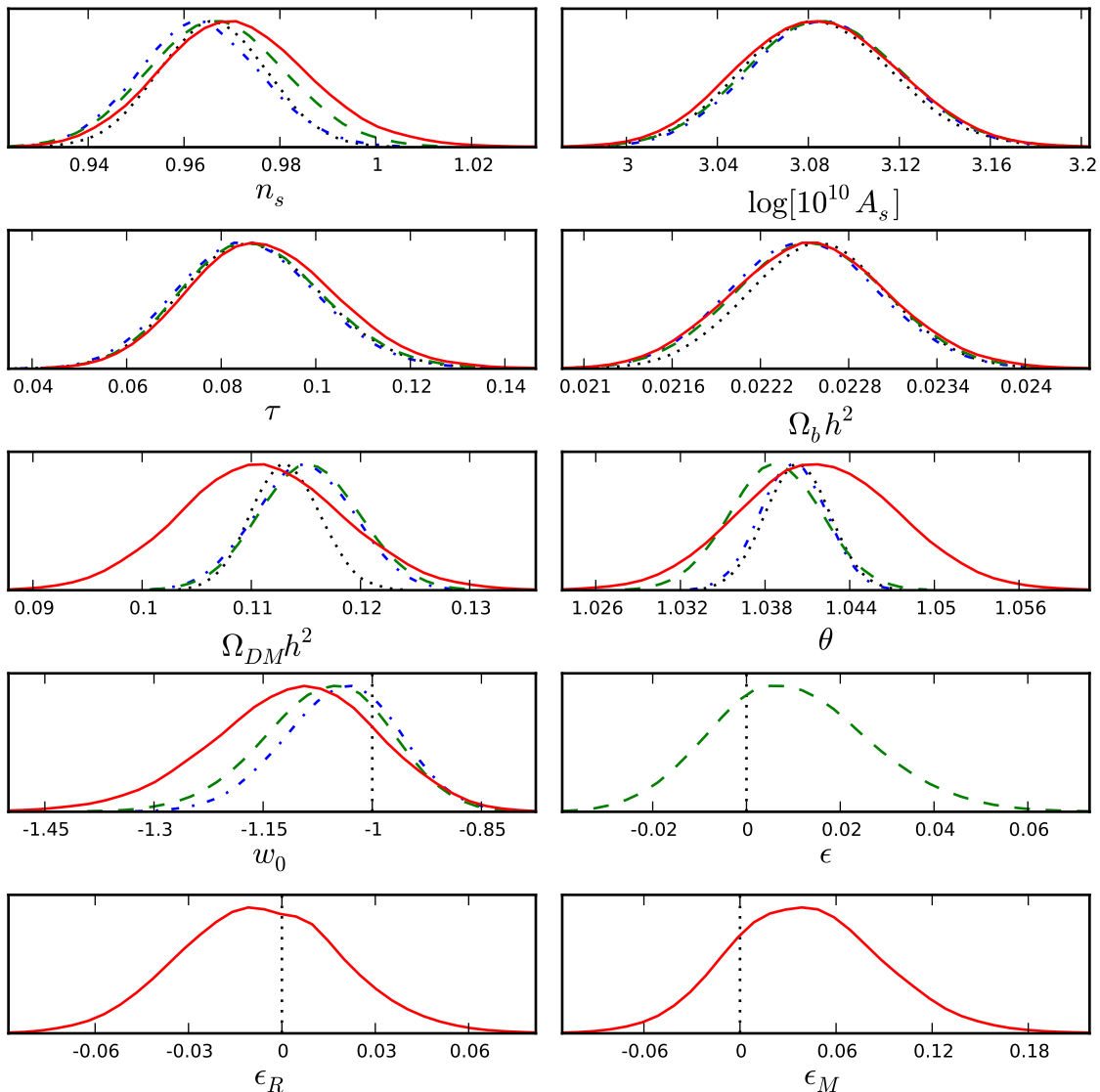


Figure 6: Posterior probabilities after marginalization over the other parameters: Λ CDM (dotted black), w CDM (dot-dashed blue), w ZCDM1 (dashed green) and w ZCDM2 (solid red); all models with unit rest frame sound speed, $\hat{c}_s = 1$.

for several values for \hat{c}_s , going as low as 10^{-5} . On the other hand, as shown in [66], the anisotropic stress in EQ models is typically small, and so we neglect it. The results are shown in fig. 9, where we compare the parameter estimation for $\hat{c}_s^2 = 1$ with that for the extreme case of $\hat{c}_s^2 = 10^{-5}$. This quantifies the dependence of the results on our assumptions on the evolution of the DE perturbations. We see that for small \hat{c}_s the value of ϵ_M is less constrained, but in any case the bounds on ϵ_M and ϵ_R remain of the same order of magnitude. The fact that the mean of the posterior of ϵ_R shifts to the right when allowing DE to cluster is in line with the tendency of the CMB data to favour an additional clustering relativistic degree of freedom, see for instance [68]. The results for $\hat{c}_s^2 = 10^{-3}$ are indistinguishable from those for $\hat{c}_s^2 = 10^{-5}$, so they are not shown. Thus, the model for

the evolution of perturbations used in the previous analysis is sufficiently general, at least at the present level of accuracy of the data. This is in agreement with the results of refs. [63, 69–71] who indeed show that, even for a small sound speed the effects are hardly detectable with current and even future data.

IV. CONCLUSIONS

In this paper we have studied a number of issues related to vacuum fluctuations in a FRW background, and the role that they could play in explaining the observed acceleration of the Universe.

First, we have studied vacuum fluctuations in a FRW background using a sharp cutoff in momentum space, and

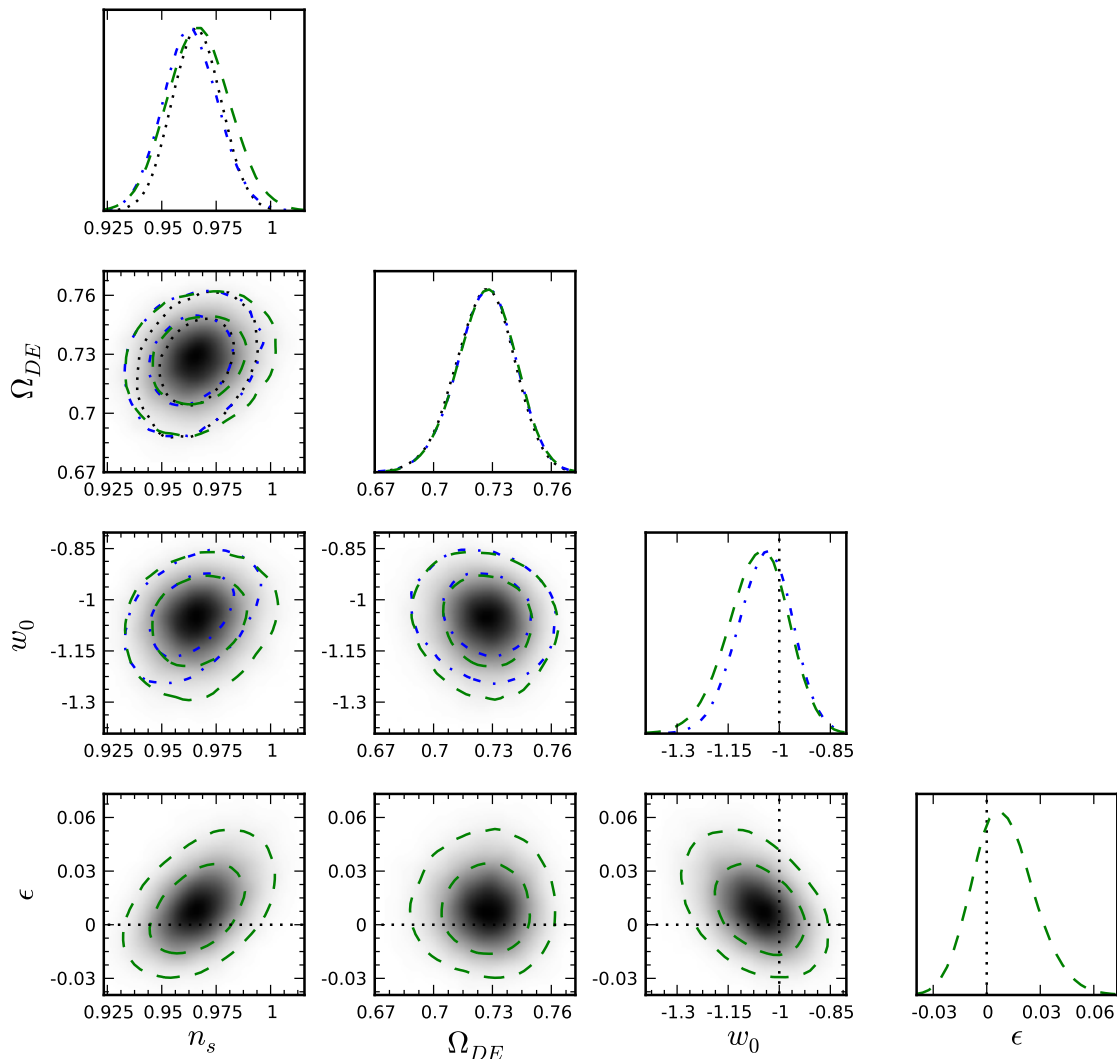


Figure 7: w ZCDM1: The marginalized likelihood contours at 1σ and 2σ are shown for all combinations of the parameters n_s , Ω_{DE} , w_0 and ϵ in w ZCDM1. We compare Λ CDM (dotted black), w CDM (dot-dashed blue) and w ZCDM1 (dashed green). The shading is the mean likelihood surface for the w ZCDM1 model. All models have unit rest frame sound speed, $\hat{c}_s = 1$.

we have explicitly shown how one can recover the same results for the physical quantities that one obtains using a covariant regularization. The basic point is that, since a momentum-space cutoff breaks general covariance, one must allow for non-covariant counter-terms. A number of apparent pathologies discussed in the literature then disappear. In fact they are seen to be a consequence of a certain confusion between the bare and the renormalized $\langle 0|T_{\mu\nu}|0\rangle$. While the bare energy density and pressure do not satisfy energy-momentum conservation, we can renormalize the theory so that the renormalized quantities do satisfy it. We have also clarified some confusion about the EOS parameter of the term proportional to $H^2(t)$ in the vacuum energy density, for which general covariance requires $w = w_{\text{tot}}(t)$ rather than $w = -1$.

Further, we have proposed an alternative to the standard renormalization of vacuum energy, in which the

Λ_c^4 divergence responsible for the cosmological constant problem is eliminated at the classical level, following a procedure that is already part of the standard tenets of classical general relativity and is used to define the ADM mass of a space-time. The crucial point here is to realize that, even if it is true that in GR the choice of the zero of the energy is not arbitrary, the correct choice is still non-trivial, and requires appropriate subtractions to be performed. We have seen that this point is further supported by the AdS/CFT correspondence, where the subtraction on the gravitational side corresponds to a renormalization of UV divergences on the CFT side. This subtraction scheme provides a possible solution to the “old” cosmological problem, namely why the vacuum energy density is not of order M_{Pl}^4 nor of order M^4 , where M is larger than the TeV scale, up to which quantum field theory is successfully applied.

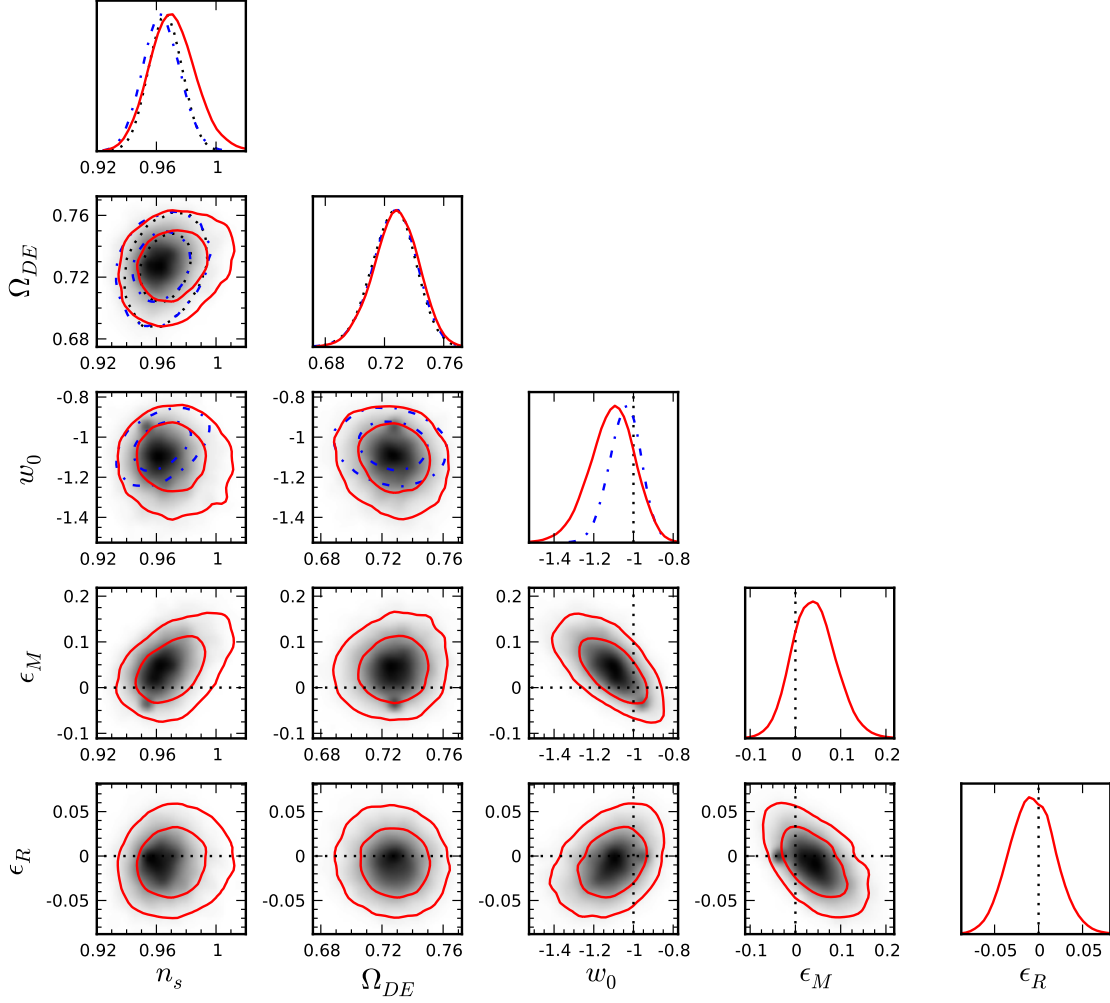


Figure 8: $wZCDM2$: The marginalized likelihood contours at 1σ and 2σ are shown for all combinations of the parameters n_s , Ω_{DE} , w_0 , ϵ_R and ϵ_M . Red solid lines are for $wZCDM2$ and blue dot-dashed lines are for $wCDM$. The shading is the mean likelihood surface for the $wZCDM2$ model. All models have unit rest frame sound speed, $\hat{c}_s = 1$.

We then turned to the contribution to vacuum fluctuations due to an ultra-light scalar field with mass $m < H_0$. We have seen that, in general, such a field generates a contribution of the form $f(\chi)R$ in the effective action (from which $\langle 0|T_{\mu\nu}|0\rangle$ is derived by functional differentiation). Such a term leads to cosmological models where the DE density tracks the dominant energy density component, thereby significantly alleviating the coincidence problem. We have performed a detailed comparison with cosmological observations of models where the DE density has the form suggested by the above theoretical considerations. We have seen that, quite generally, we can describe these models with a two-parameter extension of $wCDM$ cosmology and, using a full likelihood analysis of CMB, BAO and SNIa data, we have presented the limits on the parameters ϵ_R and ϵ_M that encode the deviations from $wCDM$. They are consistent with their ΛCDM values, as it is also the case for w_0 in $wCDM$.

Thereby there is at present no evidence for the presence of an ultralight scalar field that couples in this way to the curvature scalar.

Acknowledgements

We thank Ramy Brustein, Claudia de Rham, Rajeev Kumar Jain, Martin Kunz, Marcos Mariño, Alberto Nicolis, Valeria Pettorino, Massimo Porrati, Riccardo Rattazzi, Antonio Riotto and Misha Shaposhnikov for very useful discussions. Our work is supported by the Fonds National Suisse.

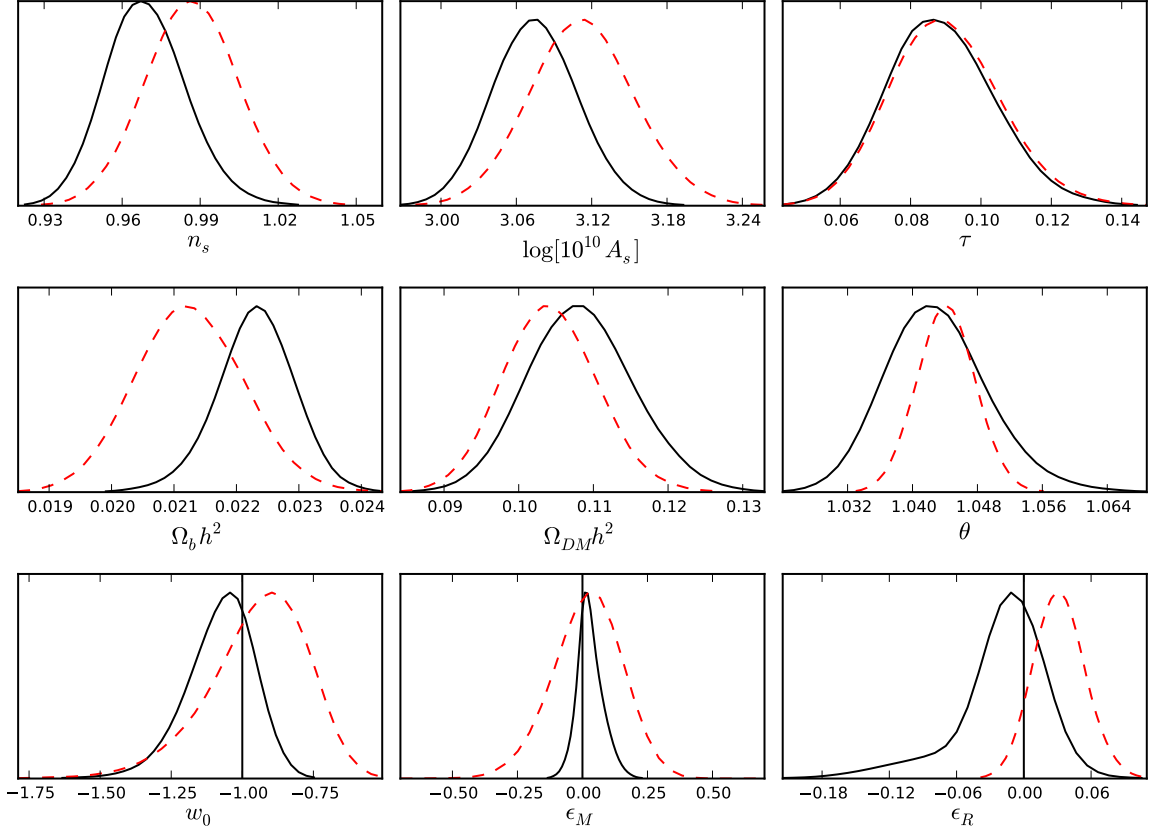


Figure 9: Posterior probabilities after marginalization over the other parameters for $w\text{ZCDM2}$ with: $\hat{c}_s^2 = 1$ (solid black) and $\hat{c}_s^2 = 10^{-5}$ (dashed red). We note that for a small \hat{c}_s , the value of ϵ_M is less constrained, a positive value is slightly preferred for ϵ_R and that w_0 is less negative.

Appendix A: Non-covariant counter-term for the $H^2(t)\Lambda_c^2$ divergence

Considering again for illustration a massless minimally-coupled scalar field, the bare energy density and pressure, as far as the term $\propto H^2(t)\Lambda_c^2$ are concerned, can be written as

$$\rho_{\text{bare}}(\Lambda_c) = \frac{H^2(t)\Lambda_c^2}{16\pi^2}, \quad (\text{A1})$$

$$p_{\text{bare}}(\Lambda_c) = \left[w_{\text{tot}}(t) + \frac{2}{3} \right] \frac{H^2(t)\Lambda_c^2}{16\pi^2}, \quad (\text{A2})$$

where we used eq. (16). A counter-term proportional to the Einstein-Hilbert action,

$$S_C = C(\Lambda_c) \int d^4x \sqrt{-g} R, \quad (\text{A3})$$

produces a counter-term

$$\langle 0 | T_{\mu\nu} | 0 \rangle = -\frac{2}{\sqrt{-g}} \frac{\delta S_C}{\delta g^{\mu\nu}} = -2C(\Lambda_c) G_{\mu\nu}, \quad (\text{A4})$$

where $G_{\mu\nu}$ is the Einstein tensor. In FRW, $G_{00} = 3H^2$ and $G_{ij} = -a^2 \delta_{ij} (2\dot{H} + 3H^2) = +3a^2 \delta_{ij} w_{\text{tot}}(t) H^2(t)$,

(where we used the relation $2\dot{H} = -3[1 + w_{\text{tot}}(t)]H^2$). This gives a counter-term for the energy density

$$\rho_C = -6C(\Lambda_c) H^2(t), \quad (\text{A5})$$

and a counter-term for the pressure

$$p_C = -6C(\Lambda_c) w_{\text{tot}}(t) H^2(t). \quad (\text{A6})$$

The function $C(\Lambda_c)$ can now be chosen so to cancel the divergence $H^2\Lambda_c^2/(16\pi^2)$ in eq. (A1) simply choosing $6C(\Lambda_c) = \Lambda_c^2/(16\pi^2) + c_1$, with c_1 a finite constant which corresponds to a finite renormalization of G , and can therefore be simply set to zero. This will also automatically cancel the divergence $w_{\text{tot}}H^2(t)\Lambda_c^2/(16\pi^2)$ in eq. (A2).

We then need a second counter-term which acts only on the pressure and cancels the term $(2/3)H^2(t)\Lambda_c^2/(16\pi^2)$ in eq. (A2). To explicitly write the required non-covariant counter-term, it is convenient to work directly in a FRW space-time with scale factor $a(t)$. In order to perform the variation of the counter-term with respect to the FRW metric we introduce $g_{\mu\nu} = (-N^2(t), b^2(t)\delta_{ij})$ and perform the variation with respect to $N(t)$ and $b(t)$. Then, *after* performing the variation, we can set the lapse

function $N(t)$ equal to one, and the function $b(t)$ equal to the desired background value $a(t)$. A counter-term S_{count} in the action induces a counter-term for the energy density given by

$$\rho_{\text{count}} = -\frac{2}{\sqrt{-g}}g_{00}\frac{\delta S_{\text{count}}}{\delta g_{00}} = -\frac{1}{a^3}\left(\frac{\delta S_{\text{count}}}{\delta N}\right)_{N=1}, \quad (\text{A7})$$

while for the pressure

$$p_{\text{count}} = \frac{1}{3a^2}\left(\frac{\delta S_{\text{count}}}{\delta b}\right)_{b(t)=a(t)}. \quad (\text{A8})$$

The counter-term that cancels the divergence $(2/3)H^2(t)\Lambda_c^2/(16\pi^2)$ in eq. (A2) will necessarily be non-covariant, and can in principle depend on the function $a(t)$ both implicitly, due to the fact that the

variation with respect to $b(t)$ is evaluated in $b(t) = a(t)$, and explicitly. A counter-term that does the job is

$$\begin{aligned} S_{\text{count}} &= -\frac{\Lambda_c^2}{96\pi^2}\int d^4x\sqrt{-\tilde{g}}(\tilde{R} + 2\tilde{R}_{00})\frac{\sum_i g_{ii}}{\sum_i \tilde{g}_{ii}} \\ &= -\frac{\Lambda_c^2}{16\pi^2}\int dt\frac{\dot{a}^2}{a}b^2(t). \end{aligned} \quad (\text{A9})$$

where $\tilde{g}_{\mu\nu} = (-1, a^2\delta_{ij})$, $g_{\mu\nu} = (-N^2, b^2\delta_{ij})$ and $\tilde{R}_{\mu\nu}$ and \tilde{R} are the Ricci tensor and Ricci scalar constructed from $\tilde{g}_{\mu\nu}$. By construction S_{count} does not depend on $N(t)$, so it does not contribute to the energy density. Its contribution to the pressure, from eq. (A8), is $-\Lambda_c^2 H^2(t)/(24\pi^2)$ and therefore cancels the term $(2/3)H^2(t)\Lambda_c^2/(16\pi^2)$ in eq. (A2).

-
- [1] N. Birrell and P. Davies, *Quantum fields in curved space* (Cambridge University Press, 340 p, 1982).
 - [2] L. Parker and S. A. Fulling, Phys. Rev. **D9**, 341 (1974).
 - [3] S. Fulling, L. Parker, and B. Hu, Phys.Rev. **D10**, 3905 (1974).
 - [4] S. A. Fulling and L. Parker, Annals Phys. **87**, 176 (1974).
 - [5] P. J. E. Peebles and B. Ratra, Rev. Mod. Phys. **75**, 559 (2003), astro-ph/0207347.
 - [6] E. Keski-Vakkuri and M. S. Sloth, JCAP **0308**, 001 (2003), hep-th/0306070.
 - [7] G. Mangano, Phys.Rev. **D82**, 043519 (2010), 1005.2758.
 - [8] M. S. Sloth, Int. J. Mod. Phys. **D19**, 2259 (2010), 1005.3241.
 - [9] T. Padmanabhan, Class.Quant.Grav. **22**, L107 (2005), hep-th/0406060.
 - [10] M. Maggiore, Phys. Rev. **D83**, 063514 (2011), 1004.1782.
 - [11] M. Maggiore, L. Hollenstein, M. Jaccard, and E. Mitsou, Phys.Lett. **B704**, 102 (2011), 1104.3797.
 - [12] A. O. Barvinsky and G. A. Vilkovisky, Phys. Rept. **119**, 1 (1985).
 - [13] I. L. Buchbinder, S. D. Odintsov, and I. L. Shapiro, *Effective action in quantum gravity* (Institute of Physics, Bristol, UK, 1992), ISBN 0-7503-0122-8.
 - [14] V. Mukhanov and S. Winitzki, *Introduction to quantum effects in gravity* (Cambridge University Press, Cambridge, 2007).
 - [15] I. L. Shapiro, Class. Quant. Grav. **25**, 103001 (2008), 0801.0216.
 - [16] E. K. Akhmedov (2002), hep-th/0204048.
 - [17] N. Bilic, S. Domazet, and B. Guberina, Phys.Lett. **B707**, 221 (2012), 1110.2393.
 - [18] N. Bilic, Phys.Rev. **D83**, 105003 (2011), 1104.1349.
 - [19] R. L. Arnowitt, S. Deser, and C. W. Misner, in *Gravitation: an introduction to current research*, edited by L. Witten (John Wiley & Sons, New York, London, 1962), chap. 7, pp. 227–265, gr-qc/0405109.
 - [20] E. Poisson, *A Relativist's Toolkit. The Mathematics of Black-Hole Mechanics* (Cambridge University Press, 2004).
 - [21] G. W. Gibbons and S. W. Hawking, Phys. Rev. **D15**, 2752 (1977).
 - [22] J. D. Brown and J. W. York, Jr., Phys. Rev. **D47**, 1407 (1993), gr-qc/9209012.
 - [23] J. D. Brown, J. Creighton, and R. B. Mann, Phys. Rev. **D50**, 6394 (1994), gr-qc/9405007.
 - [24] S. W. Hawking and G. T. Horowitz, Class. Quant. Grav. **13**, 1487 (1996), gr-qc/9501014.
 - [25] V. Balasubramanian and P. Kraus, Commun. Math. Phys. **208**, 413 (1999), hep-th/9902121.
 - [26] P. Kraus, F. Larsen, and R. Siebelink, Nucl. Phys. **B563**, 259 (1999), hep-th/9906127.
 - [27] S. R. Lau, Phys. Rev. **D60**, 104034 (1999), gr-qc/9903038.
 - [28] R. B. Mann, Phys. Rev. **D60**, 104047 (1999), hep-th/9903229.
 - [29] R. Emparan, C. V. Johnson, and R. C. Myers, Phys. Rev. **D60**, 104001 (1999), hep-th/9903238.
 - [30] R. Brustein, D. Eichler, S. Foffa, and D. H. Oaknin, Phys. Rev. **D65**, 105013 (2002), hep-th/0009063.
 - [31] K. Zhou, R.-H. Yue, Z.-Y. Yang, and D.-C. Zou (2011), 1110.0065.
 - [32] J. A. Frieman, C. T. Hill, A. Stebbins, and I. Waga, Phys.Rev.Lett. **75**, 2077 (1995), astro-ph/9505060.
 - [33] V. Sahni and S. Habib, Phys.Rev.Lett. **81**, 1766 (1998), hep-ph/9808204.
 - [34] L. Parker and A. Raval, Phys.Rev. **D60**, 063512 (1999), gr-qc/9905031.
 - [35] A. A. Starobinsky, in *Field Theory, Quantum Gravity and Strings*, edited by H. J. De Vega and N. Sanchez (Springer Verlag, 1986), pp. 107–126.
 - [36] M. Morikawa, Phys.Rev. **D42**, 1027 (1990).
 - [37] E. Calzetta and B. Hu, Phys.Rev. **D52**, 6770 (1995), gr-qc/9505046.
 - [38] J.-P. Uzan, Phys.Rev. **D59**, 123510 (1999), gr-qc/9903004.
 - [39] L. Amendola, Phys. Rev. **D60**, 043501 (1999), astro-ph/9904120.
 - [40] T. Chiba, Phys.Rev. **D60**, 083508 (1999), gr-qc/9903094.
 - [41] F. Perrotta, C. Baccigalupi, and S. Matarrese, Phys.Rev. **D61**, 023507 (2000), astro-ph/9906066.
 - [42] T. Chiba, Phys.Rev. **D64**, 103503 (2001), astro-ph/0106550.

- [43] T. Chiba, M. Siino, and M. Yamaguchi, *Phys.Rev.* **D81**, 083530 (2010), 1002.2986.
- [44] C. Wetterich, *Astron. Astrophys.* **301**, 321 (1995), hep-th/9408025.
- [45] G. Caldera-Cabral, R. Maartens, and B. M. Schaefer, *JCAP* **0907**, 027 (2009), 0905.0492.
- [46] J. Valiviita, R. Maartens, and E. Majerotto, *Mon. Not. Roy. Astron. Soc.* **402**, 2355 (2010), 0907.4987.
- [47] S. Basilakos, M. Plionis, and J. Sola, *Phys. Rev.* **D80**, 083511 (2009), 0907.4555.
- [48] J. Grande, J. Sola, S. Basilakos, and M. Plionis, *JCAP* **1108**, 007 (2011), 1103.4632.
- [49] S. Basilakos, F. Bauer, and J. Sola, *JCAP* **1201**, 050 (2012), 1109.4739.
- [50] C. M. Will, *Living Rev.Rel.* **9**, 3 (2005), gr-qc/0510072.
- [51] A. Lewis and S. Bridle, *Phys. Rev.* **D66**, 103511 (2002), astro-ph/0205436.
- [52] A. Lewis and S. Bridle, *Cosmological MonteCarlo (CosmoMC)*, publicly available Markov-Chain Monte-Carlo likelihood sampler: <http://cosmologist.info/cosmomc/>.
- [53] A. Lewis, A. Challinor, and A. Lasenby, *Astrophys. J.* **538**, 473 (2000), astro-ph/9911177.
- [54] A. Lewis and A. Challinor, *Code for Anisotropies in the Microwave Background (CAMB)*, publicly available CMB-Boltzmann code: <http://www.camb.info/www.camb.info>.
- [55] A. G. Riess et al., *Astrophys. J.* **699**, 539 (2009), 0905.0695.
- [56] O. Pisanti et al., *Comp. Phys. Commun.* **178**, 956 (2008), 0705.0290.
- [57] R. Amanullah et al., *Astrophys. J.* **716**, 712 (2010), 1004.1711.
- [58] E. Komatsu et al. (WMAP), *Astrophys. J. Suppl.* **192**, 18 (2011), 1001.4538.
- [59] C. L. Reichardt et al., *Astrophys. J.* **694**, 1200 (2009), 0801.1491.
- [60] J. L. Sievers et al., *Astrophys. J.* **660**, 976 (2007), astro-ph/0509203.
- [61] B. A. Reid et al. (SDSS), *Mon. Not. Roy. Astron. Soc.* **401**, 2148 (2010), 0907.1660.
- [62] J.-Q. Xia and M. Viel, *JCAP* **0904**, 002 (2009), 0901.0605.
- [63] R. de Putter, D. Huterer, and E. V. Linder, *Phys. Rev.* **D81**, 103513 (2010), 1002.1311.
- [64] C. L. Reichardt, R. de Putter, O. Zahn, and Z. Hou (2011), 1110.5328.
- [65] M. Doran and G. Robbers, *JCAP* **0606**, 026 (2006), astro-ph/0601544.
- [66] F. Perrotta and C. Baccigalupi, *Phys.Rev.* **D65**, 123505 (2002), astro-ph/0201335.
- [67] V. Pettorino and C. Baccigalupi, *Phys.Rev.* **D77**, 103003 (2008), 0802.1086.
- [68] E. Calabrese, D. Huterer, E. V. Linder, A. Melchiorri, and L. Pagano, *Phys.Rev.* **D83**, 123504 (2011), 1103.4132.
- [69] D. Sapone, M. Kunz, and M. Kunz, *Phys.Rev.* **D80**, 083519 (2009), 0909.0007.
- [70] G. Ballesteros and J. Lesgourgues, *JCAP* **1010**, 014 (2010), 1004.5509.
- [71] D. Sapone, M. Kunz, and L. Amendola, *Phys.Rev.* **D82**, 103535 (2010), 1007.2188.